Posttranskriptionelle Regulation des Retrotransposons TRE5-A in *Dictyostelium discoideum*

Dissertation
zur Erlangung des akademischen Grades
„doctor rerum naturalium“ (Dr. rer. nat.)

vorgelegt dem Rat der Biologisch-Pharmazeutischen Fakultät
der Friedrich-Schiller- Universität Jena

von Diplom-Biologin Anika Schmith
geboren am 20. März 1983 in Eisenhüttenstadt

1. Gutachter: Prof. Dr. Thomas Winckler, Friedrich-Schiller-Universität Jena

2. Gutachter: Prof. Dr. Hans-Peter Saluz, Friedrich-Schiller-Universität Jena

3. Gutachter: Prof. Dr. Rupert Mutzel, Freie Universität Berlin

Tag der Disputation: 26.05.2016

Inhaltsverzeichnis

1. Zusammenfassung

TRE5-A und DIRS-1 sind Retrotransposons aus dem eukaryotischen haploiden Organismus *Dictyostelium discoideum*. Von beiden existieren sowohl *Sense*- als auch *Antisense*-Transkripte. Dies führte zur Vermutung einer posttranskriptionellen Regulation beider Retrotransposons durch zelleigene RNA-Interferenz-Prozesse. Für DIRS-1 konnte in dieser Arbeit gezeigt werden, dass die RNA-abhängige RNA-Polymerase RrpC als Komponente des RNAi-Systems in *D. discoideum* für die posttranskriptionelle Regulation von DIRS-1 benötigt wird. RrpC verhindert dabei die Transkriptakkumulation von DIRS-1 und somit dessen Retrotransposition. Ebenfalls ist RrpC für die Synthese der DIRS-1 siRNAs notwendig, da im entsprechenden *rrpC-Knockout*-Stamm eine signifikante Reduzierung der siRNAs von DIRS-1 beobachtet werden konnte.

Frühere Arbeiten konnten den Wirtsfaktor C-Modul-bindender Faktor A (CbfA) aus *D. discoideum* als Transkriptionsregulator identifizieren, der *in vitro* an das C-Modul von TRE5-A bindet. In dieser Arbeit konnte gezeigt werden, dass eine reduzierte Expression von CbfA im Stamm JH.D, mit einer CbfA-Expression von nur noch 5 %, zu einem drastischen Abfall der TRE5-A-Transkriptmenge und -Retrotransposition im Vergleich zum Wildtyp führt. Die reduzierte Transkriptmenge betrifft dabei sowohl die *Sense*- als auch die *Antisense*-Transkripte von TRE5-A, wobei eine direkte Bindung von CbfA an die entsprechenden Promotorregionen von TRE5-A ausgeschlossen werden konnte. Den verschiedenen Domänen von CbfA scheinen dabei unterschiedliche Aufgaben zuteilzuwerden, da die C-terminale Domäne (CTD) von CbfA die Transkriptakkumulation, aber nicht die Retrotranspositionsfrequenz von TRE5-A beeinflussen kann.

Die in dieser Arbeit mit Wildtyp und JH.D durchgeführten vergleichenden RNA-Seq-Experimente zur weiteren Funktionsaufklärung von CbfA zeigten eine Überexpression der Argonaut-Proteine AgnC und AgnE in der CbfA-Mutante. Dies bestätigte die bereits vorhandene Vermutung, dass RNAi-Prozesse an der TRE5-A-Regulation beteiligt sein könnten. Zur Verfügung stehende *agnC*$^-$ und *agnE*$^-$-Stämme zeigten eine TRE5-A-Überexpression, was auf eine direkte Funktion der beiden Argonaut-Proteine in der TRE5-A-Regulation schließen lässt. Um *agnC* beziehungsweise *agnE* im CbfA-Wildtyp-Hintergrund überexprimieren zu können, wurden Genaktivierungsmutanten (GA) von AgnC und AgnE erzeugt. Diese Stämme wurden auf ihren Einfluss auf die TRE5-A-Expression und -Retrotransposition hin charakterisiert. Die Überexpression von *agnC* oder *agnE* führte zu einem Abfall der TRE5-A-Transkriptmenge und der GA-AgnC-Stamm zeigte eine signifikante Reduktion der Retrotranspositionsfrequenz von TRE5-A. Dies lässt auf eine

posttranskriptionelle Regulation von TRE5-A schließen. CbfA jedoch scheint diese Regulation zu supprimieren, wodurch die TRE5-A-Retrotranspositionsaktivität im Genom erhalten bleibt. Dies wiederum könnte positive Effekte auf die Genomevolution von *D. discoideum* haben.

Die in dieser Arbeit durchgeführten phylogenetischen Analysen von CbfA innerhalb und außerhalb der sozialen Amöben ermöglichten die Darstellung einer neuen Proteinfamilie: Die CbfA-ähnlichen Proteine. Dabei konnte gezeigt werden, dass die Jumonji/Zinkfinger-Region (Jmj/ZF) am N-Terminus von CbfA eine konservierte Domäne ist, die womöglich an Chromatin-Modifikationsprozessen beteiligt ist. Zwischen verschiedenen Organismen ist die CTD sehr variabel. Dies deutet auf eine schnelle evolutionäre Entwicklung dieser Domäne als Ausdruck einer artspezifischen Anpassung hin. Die CbfA-Proteine der *Dictyosteliidae*, zu denen auch *D. discoideum* gehört, zeigen dagegen eine hohe Konservierung der CTD. Die CTD von *Polysphondylium pallidum*, eine mit *D. discoideum* entfernt verwandte Art, konnte in Genexpressionsstudien viele der von der *D. discoideum*-CTD beeinflussten Gene ebenfalls regulieren.

2. Summary

TRE5-A and DIRS-1 are retrotransposons of the eukaryotic haploid organism *Dictyostelium discoideum*. Both elements produce sense and antisense transcripts and may therefore be regulated posttranscriptional by RNA interference (RNAi). In this work it was shown that the RNA-dependent RNA polymerase RrpC of *D. discoideum* is required for the posttranscriptional regulation of DIRS-1. RrpC prevents the accumulation of DIRS-1 transcripts and consequently its retrotransposition. RrpC is also required for the synthesis of DIRS-1 siRNAs because a significant reduction of siRNAs of DIRS-1 had been observed in the corresponding *rrpC*-knockout.

Previous work identified the C-module-binding factor A (CbfA) as a general regulator of transcription that binds to the C-module of TRE5-A in vitro. In this work it was shown that a reduced expression of CbfA in the mutant JH.D leads to a drastic reduction of sense and antisense transcripts and the frequency of TRE5-A retrotransposition in comparison to wild-type cells. Direct effects of CbfA to the corresponding promoter regions of TRE5-A could be excluded using reporter gene assays, assuming an indirect regulatory role of CbfA. The different domains of CbfA seem to have different functions in the regulation of TRE5-A, because the C-terminal domain of CbfA can influence the accumulation of TRE5-A transcripts but not the frequency of retrotransposition.

In this work comparative gene expression studies were performed in wild-type and JH.D cells to evaluate CbfA's gene regulatory functions. In these experiments an overexpression of Argonaute proteins AgnC and AgnE in the CbfA mutant was observed. This confirmed the assumption that RNAi-mediated processes are involved in the regulation of TRE5-A. Knock-out *agnC*$^{-}$ and *agnE*$^{-}$ strains were shown to overexpress TRE5-A, whereas gene activation mutants overexpressing *agnC* and *agnE* showed suppression of both TRE5-A expression and retrotransposition. This indicates a posttranscriptional regulation of TRE5-A. CbfA seems to repress a RNAi pathway involving AgnC and AgnE, thereby supporting the amplification of TRE5-A in the *D. discoideum* population. The continuous replication of TRE5-A may have a profound impact on the genome evolution of *D. discoideum*.

Phylogenetic analyses uncovered a new family of proteins named "CbfA-like proteins". Members of this family are widespread in dictyostelids (comprising CbfA and CbfB) and filamentous fungi and contain a "carboxy-terminal jumonji" (JmjC) domain connected to two characteristic zinc finger-like motifs that together may constitute a chromatin-modifying entity. Among different dictyostelid species the carboxy-terminal domain (CTD) is highly variable despite a highly conserved "core" that exists only in CbfA proteins. Whereas the CbfA-CTDs

are conserved enough to allow *Polysphondylium pallidum* CbfA-CTD to fully adopt the functions of the *D. discoideum* ortholog, the rapid evolution of CTDs in dictyostelid CbfB proteins and fungal CbfA-like proteins may reflect a rapid evolutionary development required for species-specific adaptation.

3. Einleitung

3.1 Der Organismus *Dictyostelium discoideum*

Dictyostelium discoideum wurde Anfang des 20. Jahrhunderts in den Wäldern North Carolinas entdeckt und als Isolat NC4 bezeichnet (*Raper, 1935*). Er gehört phylogenetisch zur Klasse der *Dictyosteliidae* (Schilde und Schaap, 2013). Diese zählen zur Gruppe der *Amoebozoa* und sind mit den Tieren und Pilzen näher verwandt als den Pflanzen (Eichinger et al., 2005). Die *Amoebozoa* sind in verschiedene Gruppen unterteilt, die eine lange unabhängige evolutionäre Entwicklung von mehreren hundert Millionen Jahren aufweisen (Tekle et al., 2008). Die Analyse eines Sets von 30 orthologen Genen aller vorhandenen *Amoebozoa*-Genome führte zu der in Abbildung 1 dargestellten Topologie. Die *Myxogastria*, mit *Physarum polycephalum* als Vertreter, und die *Dictyosteliidae* bilden dabei Schwestergruppen (Glöckner und Noegel, 2013). Die Analysen von Glöckner und Noegel (2013) konnten zeigen, dass nur einige wenige hundert *Amoebozoa*-spezifische Gene existieren. Das „Kerngenom“ des letzten gemeinsamen Vorfahren aller Eukaryoten bedarf demnach nur geringfügigen Ergänzungen, um die Gruppe der *Amoebozoa* zu etablieren. Die verschiedenen Lebensstile und Lebenszyklen mit den entsprechenden Spezifikationen führten vermutlich zu den genetischen Änderungen, die zur Entstehung der verschiedenen Linien innerhalb der *Amoebozoa* führte (Glöckner und Noegel, 2013).

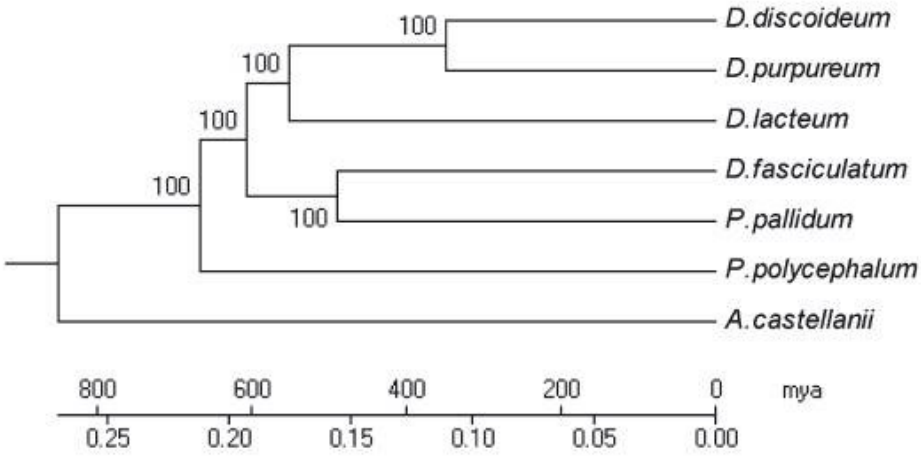


Abb. 1: Phylogenie der *Amoebozoa*. Stammbaum wurde erstellt mit Hilfe von *neighbour-joining* anhand eines Gensets von 30 Orthologen der vorhandenen *Amoebozoa*-Genome. Für die Wurzel wurde nach humanen Orthologen gesucht. A.: *Acanthamoeba*; D.: *Dictyostelium*; mya: Millionen Jahre; *P. pallidum*: *Polysphondylium*; *P. polycephalum*: *Physarum*. (Glöckner und Noegel, 2013).

Die Linie der *Dictyosteliidae* etablierte sich vor ca. 600 Millionen Jahren. Sie stellen aufgrund ihres Lebenszyklus mit Ausbildung eines multizellulären Stadiums den Übergang vom Einzeller zum Vielzeller dar. Es existieren etwa 150 *Dictyosteliidae*-Arten, die in ein phylogenetisches System eingeordnet sind (Romeralo et al., 2011). Diese können den Gattungen *Dictyostelium*, *Polysphondylium* und *Acytostelium* zugeordnet werden (Alvarez-Curto et al*.,* 2005; Schaap et al., 2006). Zunächst erfolgte die Einteilung anhand morphologischer Charakteristika. Neuere molekulare phylogenetische Analysen der Sequenzen der kleinen Untereinheit der ribosomalen rRNA und der Vergleich von α-Tubulin führten zu der Einteilung in vier Gruppen, namentlich Gruppe 1 bis 4, in welche die verschiedenen Arten von *Dictyostelium*, *Polysphondylium* und *Acytostelium* eingruppiert sind (Schaap et al., 2006). Genanalysen konnten zeigen, dass orthologe Gensets innerhalb der Arten zum großen Teil aus Elementen der allgemeinen Zellmaschinerie wie Transkription, Translation und Primärmetabolismus bestehen. Es konnten keine Art-spezifischen Genfamilien identifiziert werden, aber durchaus Art-spezifische Erweiterungen einiger Genfamilien (Glöckner und Noegel, 2013). Dabei handelt es sich meist um Funktionen, welche die interzelluläre Kommunikation betreffen (Heidel et al., 2011). Der Entwicklungszyklus der *Dictyosteliidae* scheint bereits vor circa 600 Millionen Jahren etabliert worden zu sein, da viele Gene des Entwicklungszyklus konserviert sind, was sowohl Genstruktur als auch Transkriptionsprofil betrifft (Parikh et al., 2010; Glöckner und Noegel, 2013). *D. discoideum* gehört zu den am besten erforschten Arten der *Dictyosteliidae*. Er weist eine kurze Generationszeit auf und ist unter Laborbedingungen einfach zu kultivieren. Dies und sein besonderer Lebenszyklus mit dem Wechsel vom Einzeller zu einem multizellulären Organismus machten *D. discoideum* zu einem interessanten Modellorganismus für die Erforschung von z.B. Phagozytose, Chemotaxis, Zellmigration und Signaltransduktion.

Im einzelligen Stadium ernährt sich *D. discoideum* durch Phagozytose von Bakterien. Sobald die Futterquellen versiegen, wird von den Zellen pulsartig zyklisches Adenosinmonophosphat (cAMP) sezerniert. Benachbarte Zellen reagieren, indem sie auf die Stelle der höchsten cAMP-Konzentration zuwandern und ihrerseits cAMP sezernieren, wodurch auch weiter entfernte Zellen erreicht werden (Schaap, 2011). Auf diese Weise schließen sich etwa 100.000 Zellen zu einem multizellulären Aggregat zusammen (Abbildung 2). Als Pseudoplasmodium bewegen sich die Zellen, gelenkt durch Licht und Wärme, zu einem optimalen Standort und beginnen mit der Ausbildung des Fruchtkörpers (Schaap, 2011). Dieser besteht aus vakuoliserten toten Stielzellen und einem Sporenkopf mit Sporenzellen (Bonner und Lamont, 2005). Der in dieser Arbeit verwendete Stamm AX2 konnte aus dem Isolat NC4 selektioniert werden. Er kann sowohl auf Bakterien als auch

axenisch, durch Pinozytose in bakterienfreiem Flüssigmedium, kultiviert werden (Watts und Ashworth, 1970).

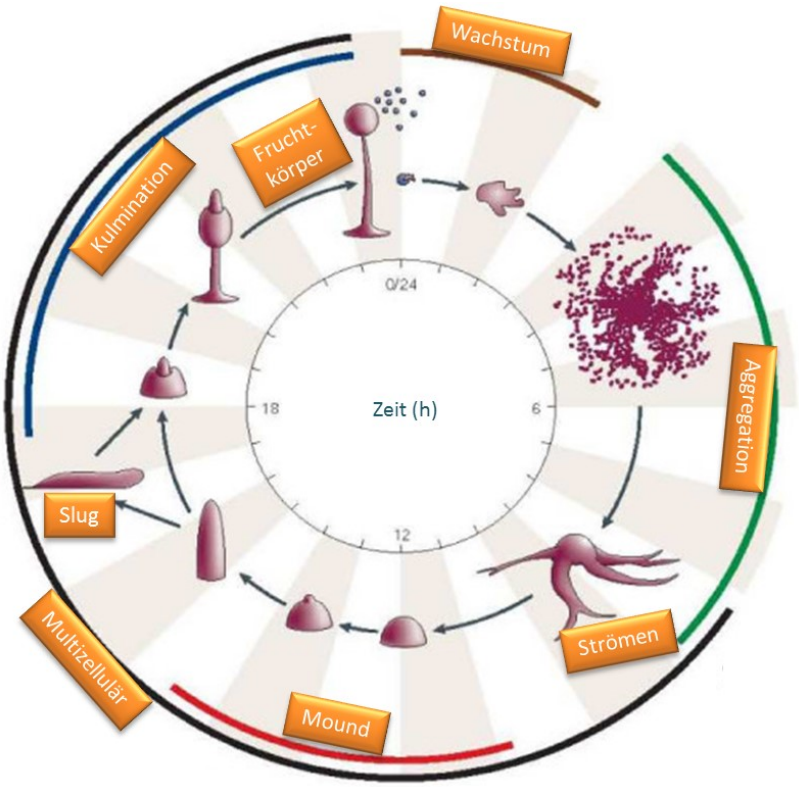


Abb. 2: Schematische Darstellung des Lebenszyklus von *Dictyostelium discoideum*. Unter günstigen Bedingungen auskeimende Sporen ernähren sich durch Phagozytose von Bakterien und vermehren sich durch Zellteilung (Wachstum). Sobald die Nahrungsquellen aufgebraucht sind, wird cAMP sezerniert, welches die Zellen zum Zentrum der höchsten cAMP-Konzentration lenkt (Aggregation). Zum Erreichen des Aggregationszentrums orientieren sich die Zellen aneinander (Strömen). Das Aggregat bildet eine schleimige Oberflächenhülle und lässt sich in Präsporen und Prästielzellen einteilen (*mound*). Der Zellkomplex wandert über das Substrat, um einen besseren Standort zu erreichen (*slug*). Ausgehend von der als *mexican hat* bezeichneten Struktur bildet sich dann der Fruchtkörper (Kulmination). Dabei schieben die schließlich absterbenden Prästielzellen die Präsporenzellen an die Stielspitze. Diese differenzieren zu Sporen aus, die von Zellulose und Glykoproteinen umhüllt sind und mehrere Jahre überdauern können. (modifiziert nach Gaudet et al., 2008).

D. discoideum hat ein haploides, 34 Mb großes Genom, das in sechs Chromosomen aufgeteilt ist. Die Chromosomen sind zwischen 4 und 9 Mb groß. *D. discoideum*-Zellen besitzen keine normalen Telomersequenzen. Stattdessen werden die Chromosomenenden von Bruchstücken des extrachromosomalen DNA-Elements gebildet, welches die ribosomalen rRNA-Gene trägt (Eichinger et al., 2005). Das vollständige 88 kb große extrachromosomale rDNA-Palindrom liegt in ungefähr 100 Kopien im Zellkern vor (Sucgang et al., 2003; Eichinger et al., 2005). Jedes Chromosom trägt an einem Ende eine Ansammlung des Retroelements DIRS (*Dictyostelium intermediate repeat sequence*), die vermutlich die Zentromere der Chromosomen bilden (Capello et al., 1984, 1985; Loomis et

al., 1995; Eichinger et al., 2005). Das Genom von *D. discoideum* weist mit 77,6 % einen hohen Gehalt an den Basen Adenin und Thymin auf. Auch besitzt das Genom eine hohe Gendichte. So findet sich statistisch alle 2,5 kb ein neues Gen. Es existieren ca. 12.500 Protein-kodierende Gene, die maximal ein bis zwei, im Mittel 129 bp große, Introns besitzen. (Eichinger et al., 2005; dictyBase). Im *D.-discoideum*-Genom finden sich 418 tRNA-Gene, von denen etwa 20 % in Paaren oder Triplets mit identischen Antikodons vorkommen (Eichinger et al., 2005; dictyBase). Es existieren außerdem ungefähr 70 weitere nicht-kodierende RNA-Gene (Aspegren et al., 2004; Hinas et al., 2006). Transposable Elemente stellen einen Anteil von ca. 10 % des Gesamtgenoms (Glöckner et al., 2001, 2002). Sie integrieren in Ansammlungen gleicher Elemente beziehungsweise im Fall der Non-LTR (*long terminal repeat*)-Retrotransposons in die Nähe von tRNA-Genen (Eichinger et al., 2005). Das haploide Genom von *D. discoideum* ist seit 2005 komplett sequenziert (Eichinger et al., 2005).

3.2 Transposable Elemente – Ein Überblick

Als Transposons oder transposable Elemente bezeichnet man DNA-Sequenzen, die befähigt sind, innerhalb des Genoms an neuen Positionen zu integrieren. Barbara McClintock hat diese „springenden Gene“ erstmals in Mais entdeckt (McClintock, 1948, 1950). Sie finden sich nahezu in allen eukaryotischen und prokaryotischen Genomen (Zaratiegiu, 2013). Ihr genomischer Anteil beträgt 3 – 20 % in Pilzen, 3 – 45 % in Tieren, 1 – 10 % in Bakterien und in Pflanzen sogar bis zu 90 % (SanMiguel et al., 1996; Mahillon und Chandler, 1998; Daboussi und Capy, 2003; Hua-Van et al., 2005; Filee et al., 2007). Transposable Elemente verbreiten sich von einer Zelle zur anderen im Wesentlichen über horizontalen Transfer und sexuelle Vererbung (Hickey, 1992; Kidwell, 1992). Aufgrund ihrer Häufigkeit und der Fähigkeit, sich im Genom frei zu bewegen, bewirken sie eine allgemeine Instabilität des Genoms (Ayarpadikannan und Kim, 2014). Ihre Insertionen können verschiedenste Effekte auf das Wirtsgenom ausüben. So können sie die Genexpression regulieren, die Rekombinationsrate erhöhen und ungleiche Cross-over hervorrufen (Ayarpadikannan und Kim, 2014). Auch können sie neue regulatorische Sequenzen schaffen, zu epigenetischen Veränderungen führen und das RNA-Editing beeinflussen (Belancio et al., 2009; Cordaux und Batzer, 2009).

Transposons werden anhand ihrer Zwischenprodukte klassifiziert: Klasse-I-Transposons, die Retrotransposons, replizieren über ein RNA-Zwischenprodukt (Finnegan, 1989). Dieses wird revers in eine DNA-Sequenz transkribiert und an einer neuen Position ins Genom integriert

(Pace und Feschotte, 2007). Klasse-II-Transposons, die DNA-Transposons, amplifizieren sich über ein DNA-Zwischenprodukt (Finnegan, 1989). DNA-Transposons besitzen endständige invertierte Sequenzwiederholungen und kodieren für eine Transposase. Diese schneidet das Element bei oder in der Nähe der Sequenzwiederholung aus dem Genom heraus und fügt es an anderer Stelle wieder ein („*cut-and-paste*“, Abbildung 3). Durch den versetzten Doppelstrangbruch entstehen kurze Sequenzverdopplungen (Finnegan, 1989; Levin und Moran, 2011) Eine Vermehrung dieser Transposons kann während der Zellteilung stattfinden, wenn die Transposition von einer Schwesterchromatide auf die andere erfolgt. Einige DNA-Transposons nutzen den Mechanismus der replikativen Transposition. Dabei verbleibt das Element an der Originalposition während eine erstellte Kopie an anderer Stelle ins Genom integriert wird (Shapiro, 1979; Johnson und Reznikoff, 1983; Parks et al., 2009; Kim, 2014). Retrotransposons verbreiten sich über einen Mechanismus, bei dem zunächst über eine *messenger*-RNA (mRNA)-Zwischenstufe eine Kopie des Elements erstellt wird, die dann an anderer Position im Genom eingefügt wird („*copy-and-paste*“) (Finnegan, 1989). Sie integrieren mit Hilfe der Integrase. Diese enthält in ihrer katalytischen Domäne das DDE-Motiv (D: Asparaginsäure, E: Glutaminsäure), welches sich auch in den Transposasen der DNA-Transposons findet und den kurzzeitigen Doppelstrangbruch ausführt (Polard und Chandler, 1995; Capy et al., 1996). Transposons unterteilen sich in autonome und nicht-autonome Elemente. Autonome Retrotransposons kodieren für die zur Reversen Transkription und Integration nötigen Proteine. Nicht-autonomen Retrotransposons fehlt mindestens eines dieser Proteine. Sie nutzen wiederum die Proteine der autonomen Elemente zur Mobilisierung (Capy, 2005). Autonome Retrotransposons untergliedern sich außerdem in LTR- (*long terminal repeats*) und Non-LTR-Retrotransposons.

LTR-Retrotransposons und die vermutlich aus ihnen entstandenen Retroviren ähneln sich in ihrer Struktur als auch in ihrem Transpositionsmechanismus (Xiong und Eickbush, 1990; Doolittle und Feng, 1992; Malik et al., 2000). Sie besitzen lange endständige Sequenzwiederholungen und kodieren für ein *gag* (*group specific antigen*)- und *pol* (Polyprotein)- Gen. Die Gag-Proteine bilden die Virus-ähnlichen Partikel (Abbildung 3). Das *pol*-Gen wiederum kodiert die zur Transposition nötigen Enzyme: Reverse Transkriptase, Ribonuklease H, Integrase und Protease (Doolittle et al., 1989). Retroviren besitzen im Unterschied zu den LTR-Retrotransposons zusätzlich ein *env*-Gen (*envelope*), welches die Infektiosität der Retroviren vermittelt und sie befähigt, die Wirtszelle zu verlassen (Bannert und Kurth, 2006). Der sich in der 5’-LTR befindende Promotor wird von der wirtseigenen RNA-Polymerase II erkannt, welche die mRNA transkribiert. Die reverse Transkription der mRNA in komplementäre DNA (cDNA) findet in Virus-ähnlichen Partikeln statt (Abbildung 3). Als Primer dient eine wirtseigene tRNA (Chapman et al., 1992). Diese bindet an eine Bindestelle unterhalb der 5‘-LTR, wodurch eine kurze Minus-Strang-DNA (*minus-strang*

strong-stop) synthetisiert wird. Diese wird zur 3‘-LTR transportiert und dient als Primer für die vollständige Minus-Strang-DNA-Synthese mit der Element-mRNA als Vorlage. Eine Polypurin-Bindestelle oberhalb der 3‘-LTR dient als Startpunkt für die Synthese einer kurzen Plus-Strang-DNA (*plus-strang strong-stop*). Diese wiederum dient an der 5‘-LTR als Primer für die Synthese der Plus-Strang-DNA, wobei die bereits synthetisierte Minus-Strang-DNA als Vorlage dient (Boeke und Corces, 1989; Boeke und Chapman, 1991). Nach Auflösen der Virus-ähnlichen Partikel erfolgt anschließend, mit Hilfe der Integrase, der Transport der linearen Doppelstrang-DNA in den Zellkern und die Integration ins Genom. Auch hier entstehen durch versetzten Doppelstrangbruch Sequenzverdopplungen (Sandmeyer *et al*., 1990; Kim *et al*., 1998).

Tyrosin-Rekombinase-Retrotransposons bilden eine eigenständige Gruppe innerhalb der Retrotransposons. Das Element DIRS (*Dictyostelium intermediate repeat* sequence) ist für diese Gruppe namensgebend. DIRS-ähnliche Elemente besitzen LTRs, jedoch sind diese im Vergleich zu den LTR-Retrotransposons nicht direkt wiederholt sondern invertiert (Goodwin und Poulter, 2001). Auch besitzen Tyrosin-Rekombinase-Retrotransposons eine interne komplementäre Region, die sowohl 5‘-LTR als auch 3‘-LTR komplementäre Sequenzen enthält (Poulter und Goodwin, 2005). Statt einer Integrase kodieren sie für eine Tyrosin-Rekombinase und bei der Integration ins Wirtsgenom entstehen keine Sequenzverdopplungen wie bei den LTR-Retrotransposons (Goodwin und Poulter, 2001). Ein möglicher Transpositionsmechanismus wird im nächsten Kapitel (3.2.1.) beschrieben.

Die Non-LTR-Retrotransposons besitzen keine LTRs. Stattdessen flankieren zwei untranslatierte Regionen (UTR) am 5‘- und 3‘-Ende des Elements (Abbildung 3) für gewöhnlich zwei offene Leserahmen (ORF). Der ORF2 kodiert die zur Retrotransposition nötigen Enzyme: Endonuklease und Reverse Transkriptase (Mathias er al., 1991; Feng et al., 1996). Für die Transposition assoziieren die ORF1- und ORF2-Proteine mit der elementeigenen mRNA (Dombroski et al., 1991; Esnault et al., 2000). Sie bilden dabei einen Ribonukleoproteinkomplex aus (Martin, 1991; Hohjoh et al., 1996; Kulpa und Moran, 2005). Die Retrotransposition findet durch einen Mechanismus statt, der als *target primed reverse transcription* (TPRT) bekannt ist (Luan et al., 1993). Während der TPRT schneidet die vom Element kodierte Endonuklease den Minus-Strang der Ziel-DNA, wodurch eine freie 3‘-Hydroxylgruppe (3’-OH) entsteht (Feng et al., 1996). Dieses freie 3’-OH-Ende dient der elementeigenen Reversen Transkriptase als Primer. Sie initiiert die cDNA-Synthese und nutzt die Element-mRNA als Vorlage (Feng et al., 1996; Cost et al., 2002; Kulpa und Moran, 2006). Dann wird vermutlich der zweite Strang der Ziel-DNA gespalten und der komplementäre DNA-Strang des Elements synthetisiert. Die Insertion erfolgt meist unter Erzeugung von Sequenzverdopplungen (Kazazian und Moran, 1998). Durch nicht vollständig

erfolgte reverse Transkription entstehen häufig 5‘-trunkierte Kopien von Non-LTR-Retrotransposons, die zur Retrotransposition nicht mehr befähigt sind (Szak *et al*., 2002).

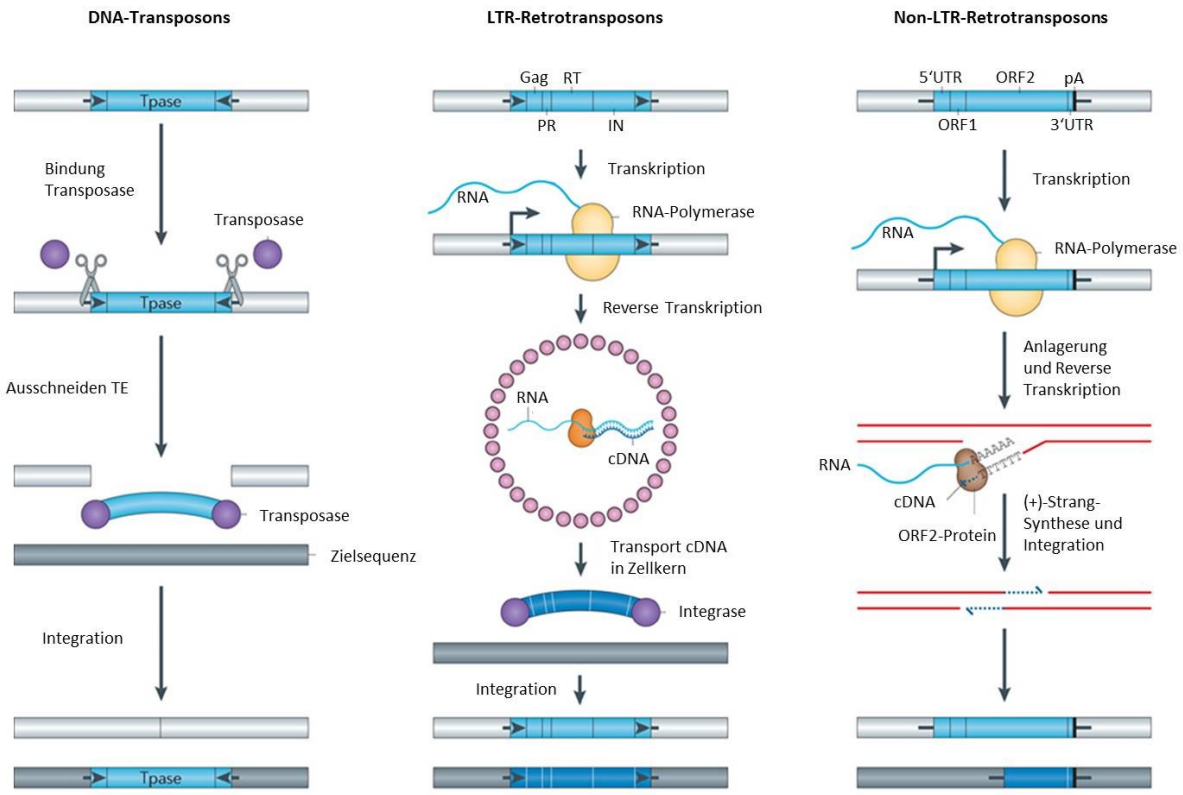


Abb. 3: Transpositionsmechanismen von Klasse-I- und Klasse-II-Transposons. DNA-Transposons kodieren für eine Transposase, welche das Element aus dem Genom herausschneidet und an anderer Stelle wieder einfügt. LTR-Retrotransposons kodieren für Gag-Proteine, Protease, Reverse Transkriptase und Integrase. Gag-Proteine bilden die Virus-ähnliche Hülle. Dort wird die mRNA in cDNA umgeschrieben. Die Integrase fügt die Kopie des Elements ins Genom ein, Non-LTR-Retrotransposons besitzen am 5‘- und 3‘-Ende untranslatierte Regionen (UTR) und meist zwei offene Leserahmen (ORF), die für die zur Mobilisierung nötigen Proteine kodieren. RNA und ORF1- und ORF2-Proteine bilden einen Ribonukleoproteinkomplex. Die Retrotransposition erfolgt durch *target primed reverse transcription.* Tpase: Transposase, TE: transposables Element, LTR: *long terminal repeat*, RT: Reverse Transkriptase, PR: Protease, IN: Integrase, UTR: untranslatierte Region, ORF: *open reading frame*, pA: poly(A). (modifiziert nach Levin und Moran, 2011).

3.2.1 Transposons in *Dictyostelium discoideum*

Die ersten transposablen repetitiven Elemente wurden in *D. discoideum* 1983 beschrieben (Rosen et al., 1983; Zuker et al., 1983). Wie in Abbildung 4 zusammenfassend dargestellt, finden sich in *D. discoideum* sowohl putative DNA-Transposons als auch Retrotransposons.

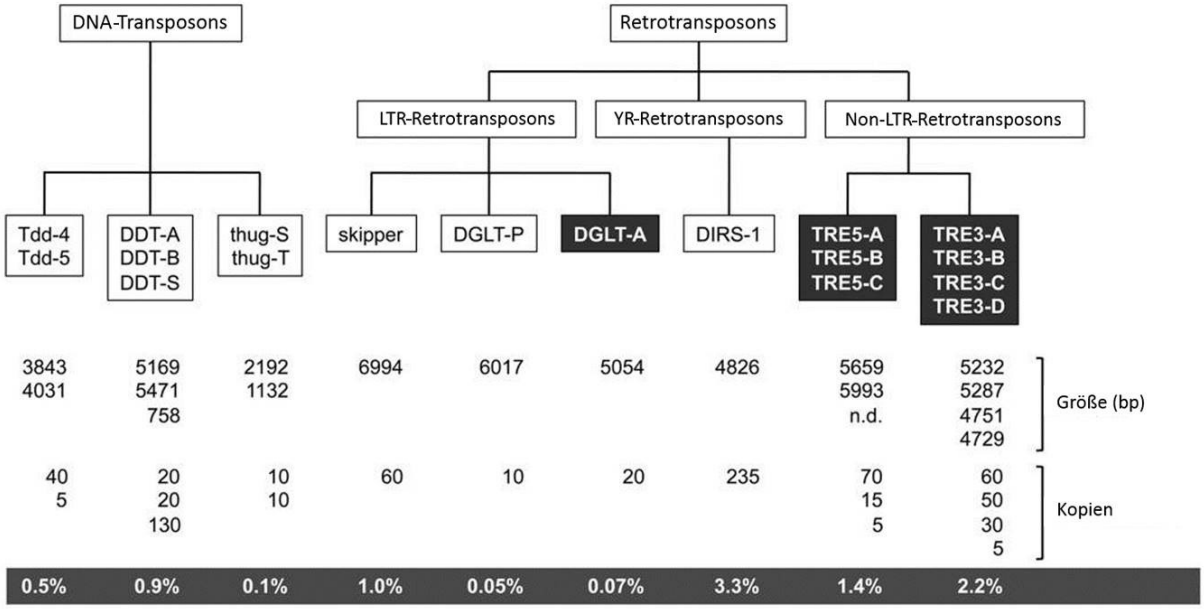


Abb. 4: Transposable Elemente in *Dictyostelium discoideum*. Die Darstellung gibt eine Übersicht über die Transposons im Genom von *D. discoideum*, ihre jeweilige Größe und ihre Kopienzahl. Die grau unterlegten Zahlen zeigen den prozentualen Anteil der Transposongruppen am Genom von *D. discoideum*. Schwarz unterlegte Transposons integrieren in der Nähe von tRNAs. YR: Tyrosin-Rekombinase, LTR: l*ong terminal repeat*, bp: Basenpaare. (modifiziert nach Winckler et al., 2011).

Die DNA-Transposons im Genom von *D. discoideum* stellen einen Anteil von circa 1,5 %. Zu ihnen gehören Tdd-4 und Tdd-5, thug-S und thug-T und die *Dictyostelium*-DNA-Transposons DDT-A, DDT-B und DDT-S. Die Einteilung zu den DNA-Transposons ist bei all diesen Elementen allerdings hypothetisch, da die kodierten Proteine keine Homologien zu bekannten Transposasen zeigen (Winckler et al., 2011). Die größte Gruppe der mobilen Elemente in *D. discoideum* bilden die Retrotransposons mit einem prozentualen Anteil von ca. 8 %. Sie untergliedern sich in LTR- und Non-LTR-Retrotransposons sowie das Tyrosin-Rekombinase-Transposon DIRS-1 (siehe Abbildung 4). Den größten Teil der LTR-Retrotransposons bildet Skipper (Glöckner et al., 2001). Es ist zwar ein stark exprimiertes Element, weist dafür allerdings eine vergleichsweise geringe Kopienzahl auf (Glöckner et al., 2001; Winckler et al., 2005). Weitere zu den LTR-Retrotransposons zählende Elemente sind DGLT-A und DGLT-P (*Dictyostelium gypsy-like transposon*). DGLT-P gilt als ausgestorbenes Element, weil kein offener Leserahmen gefunden werden konnte (Glöckner et al., 2001). DGLT-A integriert in der Nähe von tRNA-Genen (Winckler et al., 2011). Den größten Anteil im Genom von *D. discoideum* stellen die sieben Non-LTR-Retrotransposons. Sie integrieren in genau definierten Abständen zu tRNA-Genen (Glöckner et al., 2001) und werden daher als *tRNA gene-targeted retroelements* oder kurz TREs bezeichnet (Winckler et al., 2005). TREs bevorzugen entweder Integrationsstellen oberhalb (5‘) von tRNA-Genen (TRE5-A, TRE5-B und TRE5-C) oder Integrationsstellen unterhalb (3‘) von tRNA-Genen (TRE3-A,

TRE3-B, TRE3-C und TRE3-D) (Szafranski et al., 1999; Glöckner et al., 2001). Durch ihre präzisen Integrationsabstände zu tRNA-Genen integrieren TREs nicht in bereits vorhandene Kopien von TRE-Elementen, sondern im definierten Abstand zum tRNA-Gen, wodurch dort eventuell bereits vorhandene Elemente verschoben werden und sich somit Tandems bilden können (Winckler et al., 2011).

In dieser Arbeit wurden insbesondere die Expressionen der Retroelemente DIRS-1 und TRE5-A untersucht. Deswegen sollen diese Elemente im Folgenden ausführlicher dargestellt werden.

DIRS-1

Das transposable Element DIRS-1 (*Dictyostelium intermediate repeat sequence* 1) unterscheidet sich von den anderen Retrotransposons aus *D. discoideum* durch invertierte terminale Wiederholungen, die fehlende DDE-Motiv-Integrase und die fehlende Proteasedomäne (Poulter und Goodwin, 2005; Wiegand et al., 2011). DIRS-1 besitzt stattdessen eine Tyrosin-Rekombinase, welche einen katalytischen Tyrosinrest zur Rekombination zweier doppelsträngiger DNA-Stränge nutzt (Poulter und Goodwin, 2005). Im *Dictyostelium*-Genom finden sich etwa 235 Kopien von DIRS-1 (Glöckner et al., 2001). Es ist ca. 4,8 kb groß, hat lange nicht identische invertierte terminale Wiederholungen (ITR) und beginnt und endet mit dem Trinukleotid TTT (T: Thymin) (Poulter und Goodwin, 2005). DIRS-1 weist drei sich überlappende offene Leserahmen auf (siehe Abbildung 5): Der erste kodiert für das mögliche Gag-Protein, der zweite für die Tyrosin-Rekombinase und der dritte schließlich für die Reverse Transkriptase, die Ribonuklase H und eine Methyltransferase (Piednoël et al., 2011). In der Nähe der 3‘-ITR befindet sich eine interne komplementäre Region (ICR), die sowohl zur 5‘-ITR als auch zur 3‘-ITR komplementäre Sequenzen aufweist (Poulter und Goodwin, 2005).

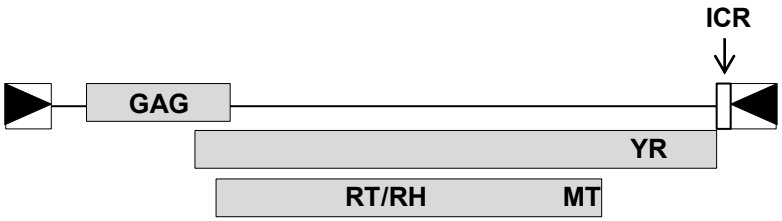


Abb. 5: Schematische Darstellung des Tyrosin-Rekombinase-Transposons DIRS-1. DIRS-1 besteht aus drei sich überlappenden Leserahmen, umrandet von invertierten terminalen Wiederholungen (ITR). Am Ende des zweiten Leserahmens vor der 3‘-terminalen Wiederholung findet sich die interne komplementäre Region (ICR). GAG: *gene specific antigen* (diese Proteine bilden den Nukleoproteinkomplex, in dem die reverse Transkription stattfindet); RT: Reverse Transkriptase; RH: Ribonuklease H; MT: Methyltransferase; YR: Tyrosin-Rekombinase. (modifiziert nach Poulter und Goodwin, 2005).

Im Vergleich zu den LTR-Retrotransposons weist DIRS-1 einen anderen Retrotranspositionsmechanismus auf. Die linke terminale Wiederholung besitzt Promotoraktivität und wird somit für die Transkription der mRNA benötigt (Cohen et al., 1984). Dabei handelt es sich um ein 4,5 kb großes Transkript mit drei sich überlappenden offenen Leserahmen (Capello et al., 1985). Die mRNA-Transkripte akkumulieren während der Entwicklung und beinhalten Teile der beiden langen terminalen Wiederholungen (Cohen et al., 1985; Wiegand et al., 2011). Durch die DIRS-1 eigene Reverse Transkriptase wird die mRNA in eine lineare cDNA transkribiert (Capello et al., 1985). Die Enden der cDNA enthalten Fragmente der beiden ITRs, die komplementär zur ICR sind. Die ICR bringt das 5'- und 3'-Ende der cDNA zusammen und dient als Vorlage, um die ITR-Sequenzen vollständig zu replizieren. Die anschließende Selbstligation erzeugt eine zirkuläre einzelsträngige DNA. Die aus dieser erzeugte doppelsträngige zirkuläre DNA inseriert abschließend mit Hilfe der Tyrosin-Rekombinase ins Genom von *D. discoideum* ohne Zielsequenzverdopplungen zu erzeugen (Capello et al., 1985; Poulter und Goodwin, 2005). Die Integration von DIRS-1 erfolgt in bereits vorhandene Kopien, so dass lange Transposon-Cluster erzeugt werden (Eichinger et al., 2005). DIRS-1 akkumuliert an jedem Ende aller sechs Chromosomen, wo diese Cluster womöglich die Zentromere von *D. discoideum* ausbilden (Eichinger et al., 2005; Glöckner und Heidel, 2009). Durch Hitzeschock und andere Stresssituationen kann die Transkription einer 900 Nukleotid langen *Antisense*-RNA hervorgerufen werden (Rosen et al., 1983; Zuker et al., 1983). Auch scheint die rechte LTR von DIRS-1 Promotoraktivität zu besitzen und ist vermutlich für die Transkription eines 4 kb großen *Antisense*-Transkripts verantwortlich (Wiegand et al., 2014). Die Regulation der DIRS-1-Expression in der Zelle erfolgt posttranskriptionell durch verschiedene Komponenten der RNA-Interferenz-Maschinerie (Boesler et al., 2014; Wiegand et al., 2014, siehe auch Abschnitt 3.4.1).

TRE5-A

Erste Studien zum Non-LTR-Retrotransposon TRE5-A wurden bereits im Jahr 1989 veröffentlicht (Marschalek et al., 1989). Es ist bekannt, dass *D.-discoideum*-Zellen eine aktive Population dieses Retroelements enthalten (Beck et al., 2002; Siol et al., 2006). TRE5-A.1 ist das komplette rund 5,7 kb große, autonome Element. Es besitzt zwei offene Leserahmen (ORF), die von untranslatierten Regionen umrandet sind, welche wiederum in „Module" eingeteilt sind (Marschalek et al., 1992; Abbildung 6). Das nicht-autonome TRE5-A.2 weist eine ORF2-Deletion auf und wird wahrscheinlich *in trans*, durch das ORF2-Protein von TRE5-A.1, mobilisiert (Beck et al., 2002; Winckler et al., 2002; Siol et al., 2006). Die Rekombination zwischen TRE5-A-Elementen führte mutmaßlich zur Deletion beider ORFs und damit zur Entstehung der „ABC-Minielemente", welche TRE5-A.3 genannt werden (Winckler et al., 2002). Wie in Abbildung 6 dargestellt, bilden das A- und B-Modul die 5'-UTR

aus und das B- und C-Modul die 3‘-UTR. Das A-Modul besitzt Promotoraktivität und führt zur Transkription der Plus-Strang RNA von TRE5-A (Schumann et al., 1994). Das B-Modul an der 5‘-UTR bestimmt den Translationsstart von ORF1, seine Funktion in der Retrotransposition ist aber noch unbekannt (Siol et al., 2011). Das C-Modul von TRE5-A stellt einen schwachen Promotor dar, der die Transkription der *Antisense*-RNA von TRE5-A vermittelt (Schumann et al., 1997). Das A-Modul ist für die generelle Retrotransposition nicht nötig, sondern kann gegen andere Promotoren ausgetauscht werden (Siol et al., 2011). Dagegen ist das C-Modul beziehungsweise eine dort gebildete Sekundärstruktur für die Transposition essentiell (Siol et al., 2011). Der ORF1 von TRE5-A kodiert für ein 51 kDa großes Protein, das möglicherweise an TRE5-A-Transkripte bindet und dabei den für die Reverse Transkription und die anschließende Integration nötigen Ribonukleoproteinkomplex mit ausbildet (Winckler et al., 2002). Der ORF2 von TRE5-A kodiert für eine Reverse Transkriptase, eine Endonuklease und eine Cystein- und Histidin-reiche Domäne (Marschalek et al., 1992).

Abb. 6: Schematische Darstellung des Non-LTR-Transposons TRE5-A. TRE5-A besteht aus zwei sich überlappenden offenen Lesrahmen, umrandet von untranslatierten Regionen. A,B,C: Module der untranslatierten Regionen; UTR: Untranslatierte Region; ORF: Offener Leserahmen (*open reading frame*); EN: Endonuklease; RT: Reverse Transkriptase; HC: Histidin- und Cystein-reiche Domäne. (modifiziert nach Winckler et al., 2005).

Wie die Transposition von TRE5-A erfolgt, ist bislang noch nicht genau bekannt. In Anlehnung an bekannte Non-LTR-Retrotranspositionsmechanismen erfolgt sie vermutlich über die sogenannte *target-primed* r*everse transcription* (Bilzer et al., 2011). Dabei bindet ORF2 vermutlich an das 3‘-Ende der Transposon-RNA und transportiert diese zum Integrationsort. Dort wird die genomische DNA gespalten und die entstandene freie 3‘-Hydroxylgruppe als Primer für die cDNA-Synthese genutzt (Feng et al., 1996). Nach erfolgter Retrotransposition sind viele der neuen Kopien von TRE5-A am 5‘-Ende deletiert. Da sie dadurch keinen funktionellen Promotor mehr besitzen, sind sie nicht mehr mobilisierbar (Beck et al., 2002; Winckler et al., 2005). Die Integration erfolgt strikt positions- und orientierungsspezifisch rund 50 bp stromaufwärts von tRNA-Genen, wobei das 5‘-Ende des integrierten Retrotransposons stets am 5‘-Ende des tRNA-Gens liegt (Marschalek et al., 1989; Hofmann et al., 1991). Diese spezifische Integration erklärt sich womöglich durch den Kontakt von TRE5-A zu den Transkriptionsfaktoren (TF) der RNA-Polymerase III, welche die Transkription der tRNA Gene vermitteln (Siol et al., 2006). Nach dem derzeitigen Modell

bindet der RNA-Polymerase-III-Transkriptionsfaktor TFIIIC zuerst an die A- und B-Box, die Promotorelemente der tRNA-Gene. Nach erfolgreicher DNA-Bindung rekrutiert TFIIIC den Transkriptions-Inintiationsfaktor TFIIIB. Das ORF1-Protein von TRE5-A interagiert dann mit einer Untereinheit des TFIIIB, dem TBP (TATA-Bindeprotein), und bestimmt somit den Integrationsort von TRE5-A (Siol et al., 2006; Chung et al., 2007; Winckler, 2013).

Beck et al. entwickelten 2002 ein Retrotranspositionssystem („TRE-Falle“, Abbildung 7), welches es ermöglicht, die Transpositionsfrequenz von TRE5-A *in vivo* zu bestimmen. Für diesen Test werden *D. discoideum*-Stämme verwendet, welche durch fehlende Expression einer Uridinmonophosphat-Synthase (UMPS) einen Defekt in der Pyrimidin-Neusynthese aufweisen. Die UMPS in *D. discoideum* wird vom *pyr5-6*-Gen kodiert und stellt ein bifunktionales Enzym dar, welches aus der Orotatphosphoribosyltransferase und der Orotidylatdecarboxylase besteht (Jacquet et al., 1988). Durch diese enzymatischen Aktivitäten kann Orotat in UMP umgebaut werden. Zellen mit defekter UMPS können kein UMP synthetisieren und sind Uracil-auxotroph (*ura*$^-$). Sie benötigen von außen zugeführtes Uracil als Vorstufe für die Bildung von Pyrimidinbasen. Gleichzeitig sind *ura*$^-$-Zellen resistent gegen das Zytostatikum 5-Fluoroorotat (5-FO): Durch die fehlende UMPS wird 5-FO nicht in 5-Fluoro-UMP umgewandelt, welches zu 5-Fluoro-Desoxy-UMP reduziert wird, einem irreversiblen Inhibitor der Thymidylat-Synthase. Im Retrotranspositionstest werden *ura*$^-$-Zellen mit einem intakten UMPS-Gen transformiert und sind somit in der Lage, UMP zu bilden. Diese Zellen sind *ura*$^+$. Das transformierte UMPS-Gen trägt in einem artifiziellen Intron ein tRNA-Gen, welches als Integrationsziel für TRE5-A-Elemente dient. In der Zelle wird das Intron inklusive des tRNA-Gens normal gespleißt, so dass funktionale UMPS entsteht. Die *ura*$^+$-Zellen werden in einem Medium ohne Zusatz von Uracil selektioniert. Sobald aber ein Transpositionsereignis am artifiziellen tRNA-Gen stattfindet, wird das UMPS-Gen durch Insertationsmutagenese inaktiviert. Die Zellen sind *ura*$^-$ und können in Medium mit Zugabe von Uracil und 5-FO selektioniert werden. Demnach entspricht jeder in 5-FO selektionierte Klon einem Sprungereignis eines endogenen TRE5-As.

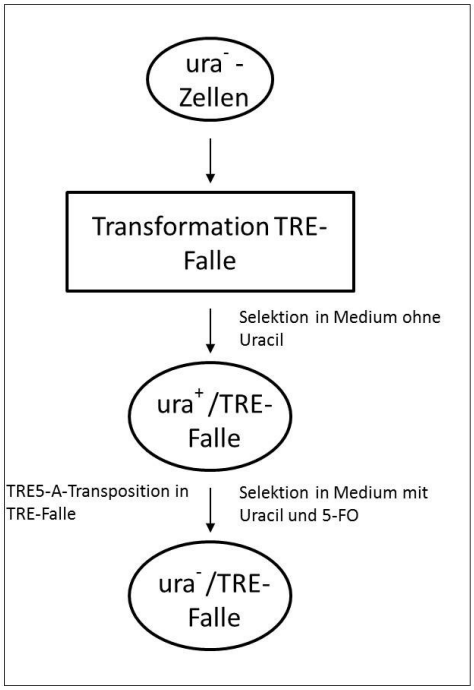


Abb. 7: Schematische Darstellung des Testsystems zur Bestimmung der *In-vivo*-Transpositionsfrequenz von TRE5-A in *D. discoideum*. Retrotranspositionssystem mit „TRE-Falle“. In 5-FO-haltigem Medium selektionierte Klone mit transformierter „TRE-Falle“ entsprechen einem Sprungereignis von TRE5-A. ura$^-$: Uracil-auxotroph; ura$^+$: Bildung Uracil; 5-FO: 5-Fluoroorotat.

3.3 Der C-Modul-bindende Faktor A

Der C-Modul-bindende Faktor A (CbfA) wurde bei der Suche nach Proteinen entdeckt, die *in vitro* an das C-Modul von TRE5-A aus *D. discoideum* binden (Geier et al., 1996). Das Gen *cbfA* befindet sich auf Chromosom 3 an der Position 2308619 bis 2312032, umfasst 3414 bp und wird von einem Intron mit einer Länge von 412 bp unterbrochen (Horn et al., 1999). Das Protein CbfA hat eine Länge von 1000 Aminosäuren mit einem Molekulargewicht von 115 kDa und besitzt drei Proteindomänen (siehe Abbildung 8). Am N-Terminus befindet sich die hoch konservierte JmjC/ZF-Region: Von Aminosäure 113 bis 280 befindet sich die Jumonji-Domäne (JmjC), die eine signifikante Ähnlichkeit zur Jumonji-Domäne des murinen Jumonji-Proteins aufweist (Takeuchi et al., 1995). Von Aminosäure 373 bis 414 und von 492 bis 550 finden sich die zwei Zinkfinger-ähnlichen Domänen (ZF). Eine Asparagin-reiche Region mit einer Länge von 217 Aminosäuren und einem Anteil an Asparagin-Resten von rund 50 % trennt die JmjC/ZF-Domäne von der C-terminalen Domäne (CTD). Die CTD hat eine Länge von 230 Aminosäuren und trägt den AT-Haken (A: Adenin; T: Thymin; Winckler et al., 2004; Lucas et al., 2009). Der AT-Haken mit seinem zentralen GRP-Motiv (G: Glycin, R: Arginin, P: Prolin) ist für die Bindung von CbfA an die AT-reiche DNA-Sequenz des C-

Moduls von TRE5-A über die kleine Furche der DNA essentiell (Horn et al., 1999). Die CTD konnte als eigenständige genregulatorische Einheit identifiziert werden (Lucas et al., 2009). Dafür wurden GFP-CTD-Fusionsproteine in sowohl Wildtyp- als auch in Cbfa-depletierten Zellen (JH.D, siehe unten) exprimiert und die allgemeine Genexpression in beiden Stämmen mit Hilfe von *Microarray*-Experimenten überprüft und verglichen. Dabei konnten viele unterschiedliche Gene identifiziert werden, die allein durch die CTD von CbfA reguliert werden können (Lucas et al., 2009). Die Funktion der anderen Domänen ist noch unklar. Die JmjC-Domäne könnte eine Chromatin modifizierende Aktivität besitzen, da JmjC-Domänen-Proteine an der Demethylierung von Arginin- und Lysin-Aminogruppen in Histonen beteiligt sind (Trewick et al., 2005; Tsukada et al., 2006). Ein entsprechender Nachweis steht für CbfA jedoch noch aus.

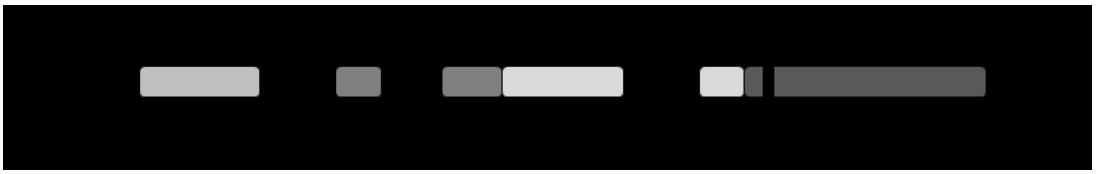


Abb. 8: Schematische Darstellung von CbfA. Die Struktur von CbfA besteht aus der JmjC/ZF-Region, die sich aus der C-terminalen Jumonji-Domäne (JmjC) und zwei Zinkfinger-Motiven (ZF) zusammensetzt, gefolgt von einer Asparagin-reichen (NRD) und der C-terminalen Domäne (CTD) mit AT-Haken und GRP-Motiv mit essentiellem zentralem Arginin zur Bindung an Adenin- und Thymin-reiche (AT)-Sequenzen. (modifiziert nach Lucas et al., 2009).

Um mehr über die Funktion des Transkriptionsregulators zu erfahren, wurde versucht, das Gen *cbfA* im haploiden Genom von *D. discoideum* durch homologe Rekombination zu zerstören. Da diese Versuche misslangen, wurde geschlussfolgert, dass es sich bei CbfA vermutlich um ein für die Zelle essentielles Protein handeln könnte (Horn et al., 1999; Winckler et al., 2001). Daher wurde versucht, Mutanten mit geringer Expression von CbfA zu generieren. In ein genomisches Fragment von *cbfA* wurde zwischen den beiden Zinkfinger-Domänen ein *Amber*-Stoppkodon (UAG) eingeführt. Anschließend wurden in den dieses Fragment tragenden Vektor ein Suppressor-tRNA-Gen und eine Blasticidin-Resistenzkassette eingefügt. Dieses Konstrukt wurde in *D. discoideum*-Zellen transformiert. Über homologe Rekombination erfolgte der *Knockin* des Konstrukts, wobei ein Teil des *cbfA*-Gens 3‘ verschoben wurde. In der entstandenen Mutante JH.D existieren zwei *cbfA*-Kopien: Eine 5‘ verkürzte, nicht-exprimierte und eine das *Amber*-Stoppkodon tragende vollständige Kopie (Winckler et al., 2001). Da *D. discoideum* nur in rund 4 % der Gene UAG als Stoppkodon nutzt, wird die Expression der $tRNA^{Glu}$ (CUA) toleriert, ohne Phänotypen zu erzeugen (Dingermann et al., 1990; Winckler et al., 2001). In der Mutante JH.D wird zu etwa

5 % das *Amber*-Stoppkodon im modifizierten *cbfA*-Gen von der Suppressor-tRNA erkannt. Dadurch kommt es zum Einbau von Glutaminsäure und der Expression eines kompletten CbfA-Proteins. Jedoch wird zu ca. 95 % die Translation am *Amber*-Stoppkodon abgebrochen. Der verkürzten Form von CbfA fehlt die CTD und sollte daher nicht mehr an seine Zielsequenzen binden können und somit nicht mehr funktionell sein (Winckler et al., 2001).

JH.D zeigt einen starken Entwicklungsdefekt (siehe Abbildung 9), weil Gene, die für die multizelluläre Entwicklung benötigt werden, nicht exprimiert werden können (Winckler et al., 2004). Die Ursache dafür liegt möglicherweise in der fehlenden Bindung von CbfA an den *acaA*-Promotor (Siol et al., 2006a). Durch die dann fehlende Expression der Adenylatzyklase A (ACA) sind *D.-discoideum*-Zellen nicht in der Lage, cAMP für die interzelluläre Kommunikation während der Aggregation zu bilden. Somit werden auch alle durch cAMP-Pulse induzierten, für die späte multizelluläre Entwicklung nötigen Gene nicht aktiviert (Winckler et al., 2004; Siol et al., 2006a). Dieser Aggregationsdefekt kann jeweils durch die Expression von rekombinantem CbfA, die Expression rekombinanter ACA, die Mischung von JH.D- und Wildtypzellen sowie die Gabe von cAMP-Pulsen ganz oder teilweise behoben werden (Winckler et al., 2004; Siol et al., 2006a).

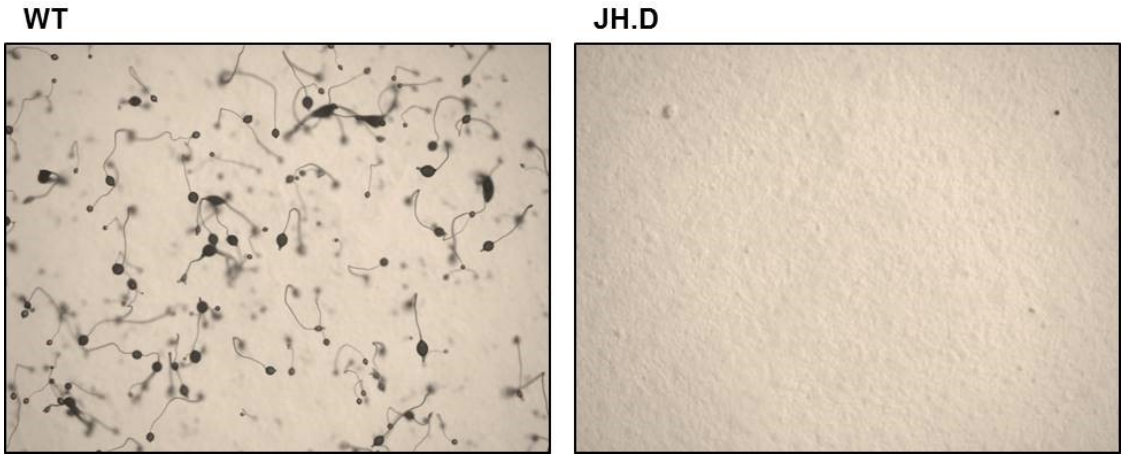


Abb. 9: Entwicklungsdefekt von JH.D. CbfA wird in JH.D nur noch zu 5 % exprimiert. JH.D ist zur multizellulären Entwicklung nicht mehr fähig. Abgebildet sind sich vollständig entwickelte Fruchtkörper von Wildtypzellen (WT) im Vergleich zur Entwicklungsmutante JH.D. (Quelle: Eigene Daten).

3.4 RNA-Interferenz

Im Jahr 1998 injizierten Fire und Mello doppelsträngige RNA (dsRNA) in *Caenorhabditis elegans* und entdeckten, dass diese dsRNA zum Abbau der komplementären mRNA in *C. elegans* führte (Fire et al., 1998). Der Prozess der Genstilllegung über dsRNA ist als RNA-Interferenz (RNAi) bekannt. RNAi ist in fast allen Eukaryoten vorhanden (Dogini et al., 2014). Dieser Prozess kann experimentell genutzt werden, um beispielsweise Genfunktionen zu identifizieren oder essentielle Gene von Pathogenen auszuschalten. Damit entwickelte sich RNAi zu einem wichtigen Werkzeug der Molekularbiologie (Dogini et al., 2014).

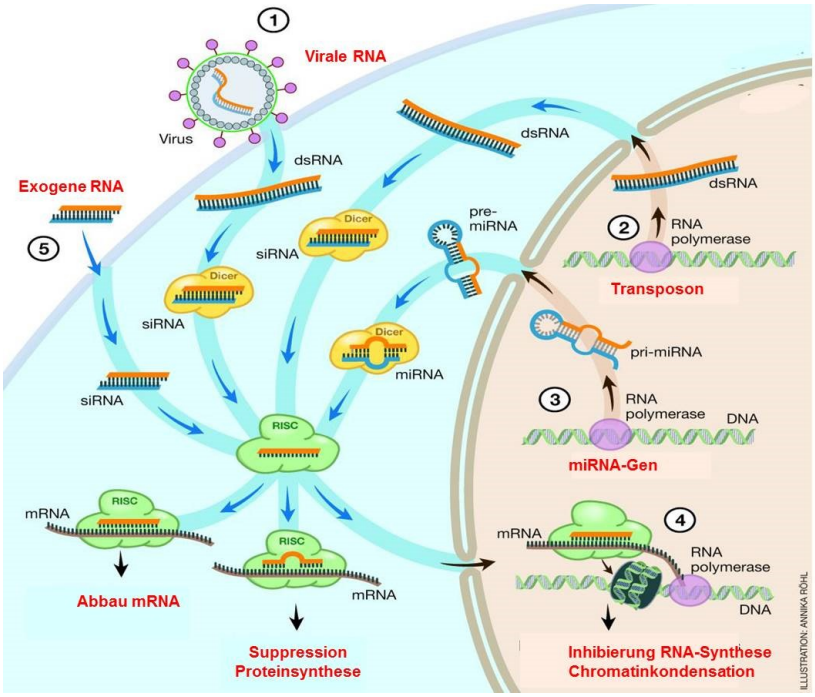


Abb. 10: Schematische Darstellung zellulärer RNAi-Prozesse. Dicer- und RISC-Komplexe sind verantwortlich für den Abbau von viraler RNA (1) und Transposontranskripten (2). Sie blockieren die Proteinsynthese durch Prozessierung und Ladung kleiner in der Zelle generierter RNAs (3) und spielen eine Rolle in der RNAi-vermittelten Inhibierung der Translation (4). Auch werden diese Komplexe genutzt, um experimentell spezifische Gene auszuschalten (5). (modifiziert nach www.nobelprize.org/nobel_prizes/medicine/laureates/2006/advanced.html).

Der generelle Ablauf der RNAi unterteilt sich in drei wesentliche Schritte: I.) Endo- oder exogene lange dsRNAs, die beispielhaft von Viren oder Transposons stammen, werden von Ribonuklease-III-ähnlichen Dicer-Enzymen in kleine sogenannte siRNAs prozessiert (Abbildung 10, Bernstein et al., 2001). II.) Diese kleinen siRNAs werden entwunden und ein Strang wird in einen Proteinkomplex geladen, der als *RNA-induced silencing complex* (RISC) bekannt ist (Abbildung 10, Tuschl et al., 1999; Hammond et al., 2000). III.) Die geladene

siRNA führt den Proteinkomplex zu komplementären Ziel-mRNAs, die vom Argonauten-Protein – als Kernkomponente von RISC – degradiert werden (Siomi und Siomi, 2009). Außerdem kann RISC die Proteinsynthese inhibieren und Chromatin modifizieren (Abbildung 10, Malone und Hannon, 2009). RNAi-Mechanismen sind ebenfalls an transkriptionellen Gen-*Silencing*-Prozessen beteiligt. Dabei rufen doppelsträngige RNA-Moleküle womöglich die Methylierung von Promotorsequenzen hervor, wodurch Promotoren, wahrscheinlich durch Chromatinkondensation (Abbildung 10), inaktiviert werden (Sijen et al., 2001). Die an RNAi beteiligten kleinen RNAs sind *small interfering* (si), *micro* (mi) und *piwi-interacting* (pi) RNAs. siRNAs mit einer Länge von 21 Nukleotiden (nt) sind an der Genomregulation und der Abwehr von Viren und Transposons beteiligt (Ghidiyal und Zamore, 2009; Malone und Hannon, 2009). Die 26 bis 33 nt langen piRNAs binden an Argonauten-Proteine der PIWI-Klade. Ihre Biogenese erfolgt Dicer-unabhängig in einem Amplifizierungsprozess, der als *ping-pong loop* bezeichnet wird. Sie sind an der Regulation der Chromatin-Beschaffenheit und der Transposonaktivität beteiligt (Ghidiyal und Zamore, 2009; Malone und Hannon, 2009; Meister, 2013). Die miRNAs mit einer Länge von 22 nt durchlaufen ebenfalls einen anderen Syntheseweg. Ihre Vorläufer-RNAs werden von der RNA-Polymerase II transkribiert und bilden doppelsträngige *Hairpin*-Moleküle (Abbildung 10) aus. Diese werden von Drosha, einem RNase-Typ-III-Enzym, erkannt, gespalten und schließlich von Dicer zu miRNA-Dimeren prozessiert (Meister, 2013).

Argonauten-Proteine sind die Kernkomponente des RISC. Die Familie der Argonauten-Proteine wird in drei Klassen unterteilt: Die AGO-Proteine, die ähnlich zu AGO1 aus *Arabidopsis thaliana* sind, des Weiteren die PIWI-Proteine, die homolog zu Piwi aus *Drosophila melanogaster* sind und schließlich die Proteine der WAGO-Klade, die in *C. elegans* anzutreffen sind (Meister, 2013). Argonauten-Proteine besitzen eine variable N-terminale Domäne, eine PAZ (*piwi-argonaute-zwille*)-, eine MID (*middle*)- und eine PIWI (*P-element induced wimpy testis*)-Domäne. Die PAZ-Domäne verankert das 3‘-Ende, die MID-Domäne das 5‘-Ende der kleinen RNA. Die N-terminale Domäne wird für die Ladung der RNA und deren Entwindung benötigt. Die PIWI-Domäne ähnelt der Ribonuklease H, agiert als Endonuklease und spaltet die Ziel-RNA (Meister, 2013).

Einige Organismen verwenden auch RNA-abhängige RNA-Polymerasen (RdRP) für die zellulären RNAi-Prozesse. RdRPs nutzen einzelsträngige RNAs als Vorlage zur Amplifizierung komplementärer RNAs. Die entstehenden dsRNAs können somit das initiale *Silencing*-Signal verstärken (Cerutti und Casas-Mollano, 2006; Wassenegger und Krezal, 2006; Maida und Masutomi, 2011). Dabei werden die dsRNAs von Dicer-Proteinen zu sekundären siRNAs prozessiert, die wiederum erneut in RISC-Komplexe geladen werden können (Girard und Hannon, 2008). RdRPs sind auch verantwortlich für einen als

Transitivität bekannten Prozess. In diesem entstehen sekundäre siRNAs, und damit *Silencing*-Signale, außerhalb von Regionen, die ursprünglich von dsRNAs getriggert wurden (Voinnet et al., 1998; Sijen et al., 2001; Vaistij et al., 2002).

3.4.1 RNAi in *Dictyostelium discoideum*

In *D. discoideum* konnten viele Komponenten identifiziert werden, die für ein funktionales RNA-Interferenz-System entscheidend sind. So finden sich zwei Dicer-ähnliche Proteine DrnA und DrnB, drei RNA-abhängige RNA-Polymerasen (RdRP) RrpA, RrpB und RrpC und die fünf PIWI-ähnlichen Argonauten-Proteine AgnA bis AgnE (Cerutti und Casas-Mollano, 2006). Auch das Vorhandensein kleiner regulatorischer RNAs wie siRNAs und miRNAs deuten auf die Präsenz eines aktiven RNAi-Systems in *D. discoideum* (Hinas et al., 2007). Bislang ist allerdings nur wenig über die Funktionen der einzelnen Komponenten bekannt. Den aktuellen Stand in der Erforschung der posttranskriptionellen RNA-Regulation in *D. discoideum* soll nun dieses Kapitel geben.

Die zwei Dicer-ähnlichen Proteine DrnA und DrnB besitzen keine N-terminale Helikase-Domäne, welche in Dicer-Proteinen anderer Organismen konserviert ist (Martens et al., 2002; Cerutti und Casa-Mollano, 2006). Dieses Motiv findet sich allerdings in den RdRPs von *D. discoideum*, möglicherweise durch einen Domänenaustausch zwischen Dicer und RdRPs (Martens et al., 2002). DrnB ist für die Prozessierung von miRNAs aus ihren Vorläufer-dsRNAs verantwortlich und auch an der Bildung sekundärer siRNAs beteiligt (Hinas und Söderbom, 2007; Avesson et al., 2012; Wiegand und Hammann, 2013).

Die drei in *D. discoideum* vorhandenen RdRP-Homolge RrpA, RrpB und RrpC besitzen am N-Terminus eine Helikase-Domäne und eine RdRP-homologe Domäne (Martens et al., 2002). RrpA und RrpB sind sich sehr ähnlich und unterscheiden sich nur in 49 von rund 2400 Aminosäuren. Das weniger konservierte RrpC zeigt dennoch Ähnlichkeit zu allen RdRPs anderer Organismen (Martens et al., 2002). Alle drei RdRPs sind an der Regulation der beiden Retrotransposons Skipper und DIRS-1 beteiligt: So führt das Ausschalten von *rrpC* zur Erhöhung der Expression von DIRS-1. In Stämmen, in denen die Expression aller drei *rrps* nicht mehr vorhanden ist, kann jeweils eine Expressionserhöhung von Skipper verzeichnet werden (Kuhlmann et al., 2005). Mittlerweile existieren K*nockout*-Stämme aller drei *rrp*-Gene, darunter ein *Triple-Knockout*, woraus geschlossen werden kann, dass diese Gene für *D. discoideum* nicht essentiell sind (Wiegand und Hammann, 2013). In der *rrpC*$^{-}$-Mutante akkumulieren sowohl *Sense*- als auch *Antisense*-Transkripte von DIRS-1 und die

Anzahl kleiner RNAs nimmt signifikant ab. Dies führt wiederum zur Translation der DIRS-1-Proteine und zur erfolgreichen Retrotransposition des Transposons (Wiegand et al., 2014).

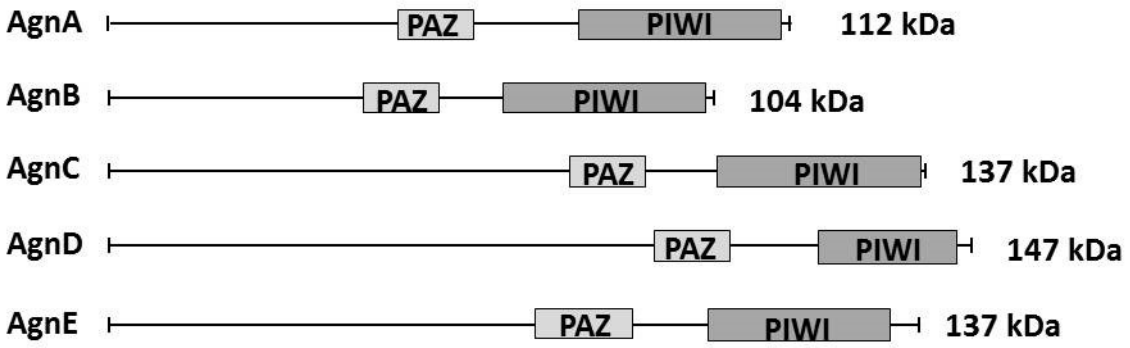


Abb. 11: Schematische Darstellung der Argonauten-Proteine von *D. discoideum*. Alle in *D. discoideum* vorhandenen Argonauten-Proteine besitzen sowohl eine RNA-bindende PAZ-Domäne als auch eine katalytische PIWI-Domäne mit Endonuklease-Aktivität. Agn: Argonauten-Protein; PAZ: *piwi-argonaute-zwille*; PIWI: *P-element induced wimpy testis*. (modifiziert nach Boesler et al., 2014).

Im Genom von *D. discoideum* kodieren fünf Gene für PIWI-ähnliche Proteine (AgnA bis E), die je eine PAZ- und eine PIWI-Domäne aufweisen (siehe Abbildung 11, Cerutti und Casas-Mollano, 2006). Über die Funktion der Argonauten-Proteine in *D. discoideum* ist bisher nicht viel bekannt. Boesler et al. konnten jedoch 2014 zeigen, dass AgnA an der DIRS-1-Regulation beteiligt ist. Ähnlich wie bei RrpC akkumulieren DIRS-1-Transkripte im *agnA*$^-$-Stamm und die Menge kleiner interferierender RNAs (siRNA) von DIRS-1 nimmt ab (Boesler et al., 2014).

4. Zielsetzung der vorliegenden Arbeit

Dictyostelium discoideum ist ein eukaryotischer Organismus aus der Gruppe der *Amoebozoa*. Trotz seines kompakten, haploiden Genoms mit einer hohen Gendichte weist *D. discoideum* eine hohe Anzahl an transposablen Elementen auf. Diese DNA-Sequenzen sind befähigt, an neuen Positionen im Genom zu integrieren. Klassifiziert werden diese Elemente in DNA-Transposons und Retrotransposons. Zu den Retrotransposons in *D. discoideum* gehören unter anderem DIRS-1 und TRE5-A. DIRS-1 akkumuliert an jeweils einem Ende der sechs Chromosomen von *D. discoideum* in bereits vorhandene DIRS-1-Kopien und trägt vermutlich zur Ausbildung der Centromere bei. TRE5-A integriert sehr spezifisch oberhalb von tRNA-Genen. Da sowohl von DIRS-1 als auch TRE5-A *Sense*- und *Antisense*-Transkripte gebildet werden, wird eine Regulation beider Elemente durch das zelleigene RNA-Interferenz-System (RNAi) vermutet. Ein Ziel dieser Arbeit war es, herauszufinden, ob TRE5-A und DIRS-1 tatsächlich von RNAi reguliert werden und die Proteine zu identifizieren, die daran beteiligt sind. Durch den Vergleich von TRE5-A und DIRS-1 sollten Unterschiede in der Regulation beider Elemente durch RNAi untersucht werden.

Vorangegangene Arbeiten mit *Microarray*-Experimenten konnten den C-Modul-bindenden Faktor A (CbfA) als generellen Transkriptionsregulator in *D. discoideum* und dessen Carboxy-terminale Domäne (CTD) als eigenständige genregulatorische Einheit identifizieren. In der Mutante JH.D, welche CbfA nur noch zu etwa 5 % im Vergleich zum Wildtyp exprimiert, konnte eine Reduktion der TRE5-A-Expression um bis zu 90 % ermittelt werden. Es sollte in dieser Arbeit daher untersucht werden, ob ein Zusammenhang zwischen RNAi-basierter Regulation von TRE5-A und der Aufrechterhaltung der TRE5-A-Expression durch CbfA besteht. Dabei sollten unter anderem RNA-Seq-Analysen der CbfA-abhängigen Genexpression von Wildtyp und JH.D miteinander verglichen und die vorhandenen *Microarray*-Daten vervollständigt werden.

Begleitend sollte eine phylogenetische Analyse der Domänenstruktur von CbfA innerhalb der sozialen Amöben weiteren Aufschluss über die Funktionen einzelner Domänen von CbfA bei der allgemeinen Genregulation und speziell bei der Regulation von Transposons geben.

5. Übersicht der Manuskripte

Manuskript 1:

The C-Module-Binding Factor Supports Amplification of TRE5-A Retrotransposons in the *Dictyostelium discoideum* Genome

Annika Bilzer, Heike Dölz, Alexander Reinhardt, **Anika Schmith**, Oliver Siol und Thomas Winckler

Publiziert in *Eukaryotic Cell*, 2011.

Inhaltsangabe: Der C-Modul-bindende Faktor A (CbfA) bindet *in vitro* an das C-Modul des Non-LTR-Retrotransposons TRE5-A. In diesem Artikel konnte gezeigt werden, dass die Akkumulation der *Sense*- als auch der *Antisense*-Transkripte von TRE5-A von CbfA abhängig ist. Auch die Retrotransposition von TRE5-A wird positiv von CbfA reguliert. Keiner der beiden internen Promotoren von TRE5-A wird dabei von CbfA reguliert. Daher wird in diesem Artikel spekuliert, dass CbfA TRE5-A indirekt reguliert, indem es Komponenten der posttranskriptionellen Gen-*Silencing*-Maschinerie supprimiert.

Eigenanteil: Der Hauptteil dieser Arbeit wurde von Annika Bilzer durchgeführt. Ich war an der Durchführung der qRT-PCR-Experimente und deren Auswertung beteiligt. Das Manuskript wurde in Zusammenarbeit von Thomas Winckler und Annika Bilzer verfasst und von mir Korrektur gelesen.

Manuskript 2:

Conserved Gene Regulatory Function of the Carboxy-Terminal Domain of Dictyostelid C-Module-Binding Factor

Anika Schmith, Marco Groth, Josephine Ratka, Sara Gatz, Thomas Spaller, Oliver Siol, Gernot Glöckner, Thomas Winckler

Publiziert in *Eukaryotic Cell*, 2013.

Inhaltsangabe: In dieser Arbeit konnte mit Hilfe von Transkriptomanalysen gezeigt werden, dass es sich bei CbfA um einen generellen Transkriptionsregulator handelt, der Genexpression positiv und negativ beeinflussen kann. Ein von CbfA stark supprimiertes Gen kodiert für das Argonauten-Protein AgnC, welches eine Komponente der RNA-Interferenz-

Maschinerie in *D. discoideum* darstellt. Des Weiteren konnte bestätigt werden, dass die C-terminale Domäne von CbfA eine eigenständige genregulatorische Einheit ist, die an der TRE5-A-Regulation beteiligt ist. Es konnte gezeigt werden, dass CbfA in der Familie der sozialen Amöben strukturell und funktionell stark konserviert wurde.

Eigenanteil: Alle Experimente zur Herstellung der Transformanten wurden von mir ausgeführt. Die Transkriptomanalysen wurden am Leibnitz-Institut für Altersforschung Jena von Marco Groth durchgeführt. Ich war an der Auswertung der Daten beteiligt und habe die qRT-PCR-Experimente zur Bestätigung der Daten durchgeführt. Die phylogenetischen Analysen erstellte ich zusammen mit Thomas Spaller. Das Manuskript wurde von Thomas Winckler und mir verfasst.

Manuskript 3:

The *Dictyostelium discoideum* RNA-dependent RNA polymerase RrpC silences the centromeric retrotransposon DIRS-1 post-transcriptionally and is required for the spreading of RNA silencing signals

Stephan Wiegand, Doreen Meier, Carsten Seehafer, Marek Malicki, Patrick Hofmann, **Anika Schmith**, Thomas Winckler, Balint Földesi, Benjamin Boesler, Wolfgang Nellen, Johan Reimegard, Max Käller, Jimmie Hällman, Olof Emanualsson, Lotta Avesson, Frederik Söderbom und Christian Hammann

Publiziert in *Nucleic Acids Research*, 2014.

Inhaltsangabe: In dieser Arbeit wurde gezeigt, dass RrpC als Komponente der RNA-Interferenz-Maschinerie in *D. discoideum* das Retrotransposon DIRS-1 posttranskriptionell supprimiert. In *rrpC-Knockout*-Stämmen akkumulieren DIRS-1-Transkripte, siRNAs von DIRS-1 verschwinden und die Transposon-Mobilität nimmt zu. Außerdem konnte dargestellt werden, dass RrpC an der Verbreitung von RNA-*Silencing*-Signalen beteiligt ist.

Eigenanteil: Der Großteil der Experimente wurde in der Arbeitsgruppe von Prof. Christian Hammann durchgeführt. Die im Artikel enthaltenen qRT-PCR-Daten zur Überprüfung der durch Northern Blots ermittelten DIRS-1-Expression in den Mutanten wurden von mir generiert und ausgewertet. Das Manuskript wurde von mir Korrektur gelesen.

Manuskript 4:

A host factor supports retrotransposition of the TRE5-A population in *Dictyostelium* cells by suppressing an Argonaute protein

Anika Schmith, Thomas Spaller, Åsa Fransson, Jonas Kjellin, Benjamin Boesler, Sandeep Ojha, Wolfgang Nellen, Christian Hammann, Fredrik Söderbom, Thomas Winckler

Publiziert in *Mobile DNA*, 2015.

Inhaltsangabe: Mit dieser Arbeit konnte belegt werden, dass CbfA die Transposition von TRE5-A positiv beeinflusst, indem es die Expression des Argonauten-Proteins AgnC inhibiert. Durch die Supprimierung von AgnC durch CbfA, können TRE5-A-Transkripte akkumulieren und als Substrat für die Retrotransposition des Elements dienen. Somit konnte gezeigt werden, dass die Amplifikation von TRE5-A in *D. discoideum* zwar von RNA-Interferenz unterdrückt wird, dies aber teilweise durch den zellulären Faktor CbfA wieder aufgehoben werden kann.

Eigenanteil: Alle experimentellen Arbeiten in diesem Paper wurden von mir durchgeführt und die erhaltenen Daten von mir analysiert. Das Manuskript wurde von Thomas Winckler und mir verfasst.

Unterschrift Betreuer

……

6. Manuskripte

6.1 Manuskript 1

The C-Module-Binding Factor Supports Amplification of TRE5-A Retrotransposons in the *Dictyostelium discoideum* Genome

---

Annika Bilzer, Heike Dölz, Alexander Reinhardt, **Anika Schmith**, Oliver Siol und Thomas Winckler

Publiziert in *Eukaryotic Cell*, 2011

EUKARYOTIC CELL, Jan. 2011, p. 81–86 Vol. 10, No. 1
1535-9778/11/$12.00 doi:10.1128/EC.00205-10

The C-Module-Binding Factor Supports Amplification of TRE5-A Retrotransposons in the *Dictyostelium discoideum* Genome[▽]

Annika Bilzer, Heike Dölz, Alexander Reinhardt, Anika Schmith, Oliver Siol, and Thomas Winckler*

School of Biology and Pharmacy, Institute of Pharmacy, Department of Pharmaceutical Biology, University of Jena, Semmelweisstrasse 10, 07743 Jena, Germany

Received 23 August 2010/Accepted 4 November 2010

Retrotransposable elements are molecular parasites that have invaded the genomes of virtually all organisms. Although retrotransposons encode essential proteins to mediate their amplification, they also require assistance by host cell-encoded machineries that perform functions such as DNA transcription and repair. The retrotransposon TRE5-A of the social amoeba *Dictyostelium discoideum* generates a notable amount of both sense and antisense RNAs, which are generated from element-internal promoters, located in the A module and the C module, respectively. We observed that TRE5-A retrotransposons depend on the C-module-binding factor (CbfA) to maintain high steady-state levels of TRE5-A transcripts and that CbfA supports the retrotransposition activity of TRE5-A elements. The carboxy-terminal domain of CbfA was found to be required and sufficient to mediate the accumulation of TRE5-A transcripts, but it did not support productive retrotransposition of TRE5-A. This result suggests different roles for CbfA protein domains in the regulation of TRE5-A retrotransposition frequency in *D. discoideum* cells. Although CbfA binds to the C module *in vitro*, the factor regulates neither C-module nor A-module promoter activity *in vivo*. We speculate that CbfA supports the amplification of TRE5-A retrotransposons by suppressing the expression of an as yet unidentified component of the cellular posttranscriptional gene silencing machinery.

Retrotransposons are ancient mobile elements that amplify in eukaryotic cells via reverse transcription of RNA intermediates (6). Because genomic integration of retrotransposon-derived cDNA is a default mechanism in the amplification of these elements, their activity causes a constant threat of insertion mutagenesis and genome instability in host genomes. This threat may be particularly important if the host cell has a haploid and compact genome, such as that of the social amoeba *Dictyostelium discoideum*, in which some 67% of nuclear DNA codes for proteins (7). *D. discoideum* mobile elements display two different regional integration preferences (reviewed in reference 28). The first group of mobile elements shows a strong bias toward inserting into preexisting copies of the same or similar elements, thereby forming large clusters in certain chromosomal regions that rarely contain protein coding capacity. A second group of retrotransposons has developed mechanisms to actively target sites in the close vicinity of tRNA genes as landmarks for "safe" integration. These elements are therefore called "tRNA gene-targeting retroelements" (TREs) (25). Because tRNA genes are scattered throughout all *D. discoideum* chromosomes and offer approximately 400 different integration sites, TREs have managed to colonize all regions of the *D. discoideum* genome.

We recently showed that TRE5-A maintains an active population in *D. discoideum* cells (2, 23). TRE5-A forms two major subpopulations, referred to as TRE5-A.1 and TRE5-A.2. These elements were formerly known as DREa and DREb (18, 19). TRE5-A.1 represents the full-length retrotransposition-competent element, whereas TRE5-A.2 has an extended ORF2 deletion and is probably mobilized in *trans* by TRE5-A.1 (2, 23). Considering the amino acid similarity of the TRE5-A ORF2 protein with other non-long terminal repeat (LTR) retrotransposons (15, 16), it is most likely that TRE5-A retrotransposition involves a coupled reverse transcription and integration process known as target-primed reverse transcription (5, 13). *In vitro* experiments suggest that TRE5-A may identify tRNA genes as integration sites by making contact between the TRE5-A-encoded ORF1 protein, as part of a preintegration complex, and the TATA-binding protein that is part of the tRNA gene-specific transcription complex TFIIIB (4, 23).

In growing *D. discoideum* cells, TRE5-A is well expressed in both directions from element-internal promoters that are recognized by RNA polymerase II (22). Plus-strand RNA [(+) RNA] is produced by a promoter located within the A module in the 5′-untranslated region of the element. Antisense RNA [(−) RNA] is generated by the C-module promoter at the 3′ end of the TRE5-A element (22). A C-module-binding factor (CbfA) was purified and characterized based on its *in vitro* binding to the C module of TRE5-A in electrophoretic mobility shift assays (9, 10, 24, 27, 29). The CbfA protein spans 1,000 amino acids and contains a "carboxy-terminal Jumonji domain" (JmjC), two zinc finger-like motifs of unknown function, an asparagine-rich domain, and a distinct carboxy-terminal domain (referred to here as CbfA-CTD). JmjC domains are thought to catalyze the oxidative demethylation of histone tails in chromatin (reviewed in references 1, 11, and 26), thereby contributing to the epigenetic control of gene transcription. It is currently unknown whether CbfA has chromatin-remodeling activity.

DNA microarray analyses have indicated that CbfA regulates at least 160 genes during the growth phase of *D. discoi-*

* Corresponding author. Mailing address: Universität Jena, Lehrstuhl für Pharmazeutische Biologie, Semmelweisstrasse 10, D-07743 Jena, Germany. Phone: 49-3641-949841. Fax: 49-3641-949842. E-mail: t.winckler@uni-jena.de.

▽ Published ahead of print on 12 November 2010.

deum (14). CbfA is also essential to initiating the multicellular life cycle of *D. discoideum* cells, probably because CbfA mutants are unable to induce the adenylyl cyclase that generates the cyclic AMP required for intercellular signaling during aggregation (27). Interestingly, DNA microarray analyses have also revealed a gene regulatory function of CbfA-CTD that does not require the remainder of the CbfA protein. In fact, some 50% of CbfA-dependent genes are regulated exclusively by CbfA-CTD (14).

Here we show that the retrotransposition of TRE5-A in *D. discoideum* cells depends on a functional CbfA protein. Whereas CbfA-CTD is sufficient to maintain high steady-state levels of both (+) RNA and (−) RNA of TRE5-A, it does not support productive retrotransposition of these TRE5-A transcripts in the absence of the full-length CbfA protein. This finding suggests different roles for CbfA protein domains in the regulation of the TRE5-A retrotransposition frequency in *D. discoideum* cells.

MATERIALS AND METHODS

***In vivo* retrotransposition assay (TRE trap assay).** JH.D[ura$^-$] cells were derived from strain JH.D (29) by prolonged selection of cells in the presence of 100 μg/ml 5-fluoroorotic acid (5-FOA) and 20 μg/ml uracil. DH1[ura$^-$] (3) and JH.D[ura$^-$] cells were cultured either in HL5 medium or in FM medium supplemented with 20 μg/ml uracil. The TRE trap assay was performed as described by Siol et al. (23). Briefly, DH1[ura$^-$] or JH.D[ura$^-$] cells were transformed with a plasmid carrying a *D. discoideum pyr5-6* gene whose reading frame was disrupted by an artificial intron. A *D. discoideum* Val^{UAC} tRNA gene was inserted into the intron as bait to attract TRE5-A integrations. These constructs are referred to as TRE^{trap} genes. TRE^{trap}-carrying plasmids were transformed into DH1[ura$^-$] or JH.D[ura$^-$] cells, and transformants were selected in FM medium in the absence of uracil. DH1[ura$^+$/TRE^{trap}] or JH.D[ura$^+$/TRE^{trap}] cells were cultured further in HL5 medium and were supertransformed either with plasmid pDXA-rCbfA, expressing nearly-full-length CbfA$^{2-998}$ (27), or with pDXA-CTD, expressing the carboxy-terminal domain of CbfA (CbfA-CTD; CbfA$^{724-998}$) (14). Stable transformants were obtained by selection in HL5 medium containing 4 μg/ml G418. To select for *de novo* retrotransposition events, 10^7 cells of the respective transformants were cultured in FM medium in the presence of 250 μg/ml 5-FOA and 30 μg/ml uracil, as described previously (2, 23). TRE^{trap}/ura$^-$ clones from five different plates were counted, and mean values ± standard deviations (SD) are presented. Each experiment was repeated at least twice.

Luciferase reporter assay. An A module of TRE5-A.1 was amplified by PCR from plasmid pB3 (18) and inserted into pGEM-T (Promega). A firefly luciferase gene, including the downstream *cotB* (*sp70*) terminator, was isolated as a HindIII/SpeI fragment from plasmid pVTL-AL (12) and then ligated into the primer-derived HindIII site downstream of the A module and the vector-derived SpeI site of pGEM-T. The resulting vector was named pGEM-A-luc. In this vector, the A module was replaced by a PCR-amplified C module to generate pGEM-C-luc. The reporter plasmids were cotransformed with pISAR (17) into *D. discoideum* AX2 and JH.D cells, and stable transformants were selected in HL5 medium containing 10 μg/ml or 7.5 μg/ml G418, respectively. Cells from plates containing at least 50 clones were pooled, and the copy number of the luciferase reporter gene was determined by quantitative PCR analysis of genomic DNA preparations. Cells from individual pools were grown in shaken cultures in HL5 medium supplemented with G418. Logarithmically growing cells were washed and frozen as cell pellets of 2×10^7 cells at −80°C until further use. Frozen cells were lysed by adding 200 μl of phosphate-buffered saline (PBS). Aliquots corresponding to 2×10^6 cells were measured using the Promega Bright-Glo luciferase assay system. Values are given as arbitrary units standardized to 100 copies of the luciferase gene per cell. For each reporter construct, 4 to 6 independent pools were tested, and the average values ± SD are presented.

Quantitative PCR. Real-time reverse transcription-PCR (RT-PCR) was performed as described previously (14). Logarithmically growing *D. discoideum* cells were washed in 17 mM phosphate buffer (pH 6.2) and stored as pellets of 2×10^7 cells at −80°C until further use. Total RNA was prepared from frozen cells by use of a Qiagen RNeasy Mini kit according to the provided manual. cDNA was synthesized by reverse transcription of 500 ng of total RNA, using an oligodesoxythymidine primer and a Qiagen Omniscript RT kit. Real-time PCR signals were standardized for expression of the gene encoding glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (*gpdA*; Dictybase entry DDB0185087) (8). The *gpdA* gene was amplified with primers gpdA-01 and gpdA-02, yielding a 247-bp PCR product from genomic DNA and a 156-bp fragment from cDNA. Thus, amplification of *gpdA* was also suitable for determining genomic DNA contamination in cDNA preparations by conventional RT-PCR prior to real-time RT-PCR runs. Real-time amplification was carried out using Stratagene Brilliant SYBR green QPCR master mix on a Stratagene Mx3000P instrument. Differential expression was calculated by the method of Pfaffl (21), using GAPDH as a reference. Results for three biological replicates are indicated (mean ± SD). TRE5-A transcript levels were determined with primers specific for the ORF1 sequence (Rep-108, 5′-GTCATAAACATCAATCCGAACCAGAC-3′; and Rep-109, 5′-GTTAGATTGTCTAGTTCAATGATAGTGTC-3′). Expression of genes encoding RNA-dependent RNA polymerases was determined using the following primers: rrpA-01, 5′-GAACGTCAAGAACTTGGTAAATTGTATCA-3′; rrpA-02, 5′-TAACCTACAGTTTGTAACCGAATGTTTAC-3′; rrpB-01, 5′-GAACGTCAAGAACTTGGTAAAATGTATAA-3′; rrpB-02, 5′-GTGGATAACCTTTAGTTTTTAACCAAAC-3′; rrpC-01, 5′-GGTGTTTATAGTAAAAAAGAATCATTC-3′; and rrpC-02, 5′-CAACTATCCAAGAATTTATGAACATTTAC-3′.

Northern blots. Agarose gel electrophoresis and blotting of total RNA were performed as described previously (27). Strand-specific DNA probes to detect TRE5-A (+) RNA and (−) RNA were prepared by cloning the BglII restriction fragment covering nucleotide positions 4610 to 5604 of TRE5-A.1 from plasmid pB3 (18) into pGEM7Zf(−) (Promega). Strand-specific radiolabeling of TRE5-A RNA was then achieved by primer extension in the presence of [α-^{32}P]dATP, using a *Taq* Cyclist DNA sequencing kit from Stratagene (27), with vector-specific primers at both ends of the cloned insert.

RESULTS

Accumulation of TRE5-A transcripts in *D. discoideum* cells requires CbfA. Although CbfA is required to regulate the expression of a multitude of genes in the *D. discoideum* genome (14), its original isolation as a C-module-binding factor suggested an additional role for the factor in the regulation of TRE5-A transcript levels and/or retrotransposition activity. To evaluate whether CbfA has an influence on the steady-state transcript levels of TRE5-A, we conducted quantitative RT-PCR on poly(dT)-primed cDNA preparations. Considering that retrotransposon activity is a source of genome instability, we expected that the function of CbfA might be to suppress the accumulation of high levels of TRE5-A transcripts, the prerequisite for retrotransposition. To our surprise, we observed that the depletion of CbfA from *D. discoideum* cells in the mutant strain JH.D did not lead to stabilization, but rather to a loss of TRE5-A transcripts. A reduction in TRE5-A expression was determined, by a factor of 5.3 ± 1.5 ($n = 7$), compared with that of the parent strain AX2.

To evaluate whether the accumulation of both (+) RNA and (−) RNA TRE5-A transcripts was affected in CbfA-depleted cells, we prepared Northern blots of total RNA from *D. discoideum* cells and used strand-specifically radiolabeled DNA probes to visualize (+) RNA and (−) RNA. To increase the sensitivity of the assay, we treated cells in parallel cultures with the respiratory chain blocker antimycin A, which is known to increase the amount of TRE5-A transcripts in growing *D. discoideum* cells (22). As shown in Fig. 1, the levels of both (+) RNA and (−) RNA were drastically reduced in the absence of a functional CbfA protein.

To determine if the observed loss of TRE5-A transcripts in the CbfA mutant was in fact related to the depletion of CbfA, we performed complementation studies using two plasmid-borne CbfA variants. The first was a nearly-full-length CbfA$^{2-998}$ protein expressed from the strong *act15* promoter.

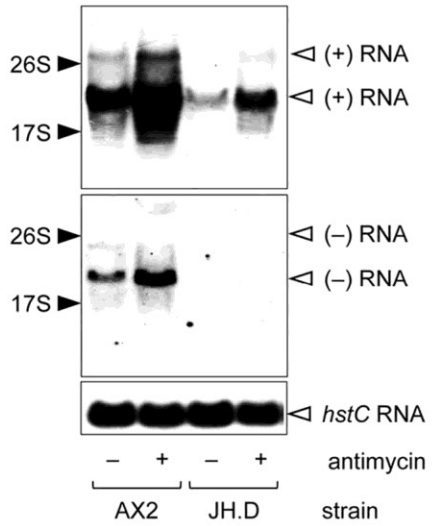


FIG. 1. CbfA is involved in regulation of TRE5-A steady-state RNA levels. AX2 and JH.D cells were grown to late logarithmic phase in the absence or presence of 80 μM antimycin A as described previously (22). Total RNA was prepared, separated by agarose gel electrophoresis, and blotted. The filter was first hybridized with a (−) RNA strand-specific probe, then stripped and hybridized with a (+) RNA strand-specific TRE5-A DNA probe, and finally stripped again and hybridized to histone H3 (*hstC*) RNA as a loading control. Size markers to the left refer to the migration positions of 26S and 17S rRNAs (ca. 4.1 and 1.9 kb, respectively).

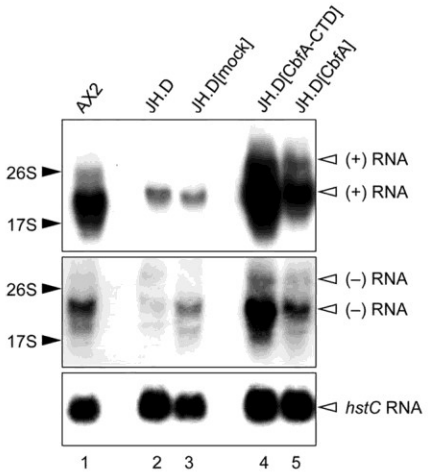


FIG. 2. Ectopic expression of CbfA in a CbfA-depleted mutant rescues TRE5-A expression. A Northern blot was prepared with total RNAs from the indicated strains: lane 1, AX2 cells; lane 2, untransformed JH.D cells; lane 3, JH.D cells transformed with empty expression vector; lane 4, JH.D cells expressing CbfA-CTD; and lane 5, JH.D cells expressing full-length CbfA. The filter was hybridized successively with (−) RNA strand-specific and (+) RNA strand-specific TRE5-A DNA probes, followed by a probe detecting histone H3 (*hstC*) RNA as a loading control. Size markers to the left refer to the migration positions of 26S and 17S rRNAs (ca. 4.1 and 1.9 kb, respectively).

The CbfA$^{2\text{-}998}$ protein proved to be functional in restoring the aggregation phenotype of JH.D cells (27). The second was the carboxy-terminal domain of CbfA (CbfA$^{724\text{-}998}$; CbfA-CTD), also expressed from the *act15* promoter, which is known to act as an independent gene regulatory entity for about half of all CbfA-regulated genes (14). When either full-length CbfA or CbfA-CTD was produced in the JH.D background, complete reversion of both TRE5-A (+) RNA and (−) RNA expression was observed (Fig. 2).

A possible explanation for the simultaneous loss of TRE5-A (+) RNA and (−) RNA in the CbfA mutant could be direct positive regulation of the two responsible TRE5-A promoters, the A module and the C module, by CbfA. To test this possibility, we inserted the A module and the C module upstream of a luciferase reporter gene in a *D. discoideum* expression vector. AX2 and JH.D cells were transformed with these vectors, and luciferase reporter gene activity was determined. Both promoters mediated expression of luciferase, with reporter activity above the background, but we did not observe significant differences in A-module or C-module promoter activity in the CbfA-depleted mutant JH.D cells in comparison to AX2 cells (Fig. 3).

CbfA increases TRE5-A retrotransposition activity in *D. discoideum* cells. The results described above indicate that CbfA is a host factor for TRE5-A expression in growing *D. discoideum* cells. We wanted to determine whether CbfA increased not only the steady-state levels of TRE5-A-derived transcripts but also the retrotransposition frequency of the element. First, we measured the frequency of *de novo* TRE5-A retrotransposition events in the CbfA-depleted mutant strain JH.D, using the previously established "TRE trap" assay. This assay reliably reflects the natural retrotransposition activity of the TRE5-A population in the *D. discoideum* genome (2, 23). Briefly, the TRE trap assay consists of a cloned *D. discoideum* UMP synthase gene (*pyr5-6*) that contains an artificial intron (Fig. 4). Into this intron, we placed a tRNA gene that served as a target for integration of TRE5-A elements, thus disrupting the entire *pyr5-6* gene. When such a TRE^{trap} gene was transformed into the ura$^-$ *D. discoideum* strain DH1, the TRE^{trap} plasmid readily complemented the ura$^-$ phenotype and converted the cells to a ura$^+$ phenotype. Such cells, however, were

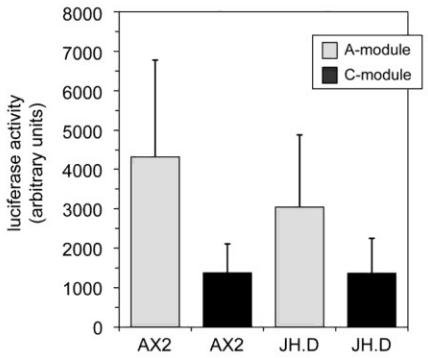


FIG. 3. Activities of the A module and the C module in the CbfA mutant strain JH.D. AX2 and JH.D cells were transformed with plasmid pGEM-A-luc (gray bars) or pGEM-C-luc (black bars) and with pISAR. After G418 selection, clones were pooled, and the average copy numbers of the expression plasmids were determined by quantitative PCR. Only pools with average plasmid copy numbers between 25 and 80 were selected for luciferase reporter assays. The values resulted from 4 to 6 pools per transformation and were normalized for 100 plasmid copies. Background activities of AX2 cells and JH.D cells transformed only with pISAR were 9 ± 10 and 17 ± 17 arbitrary units, respectively.

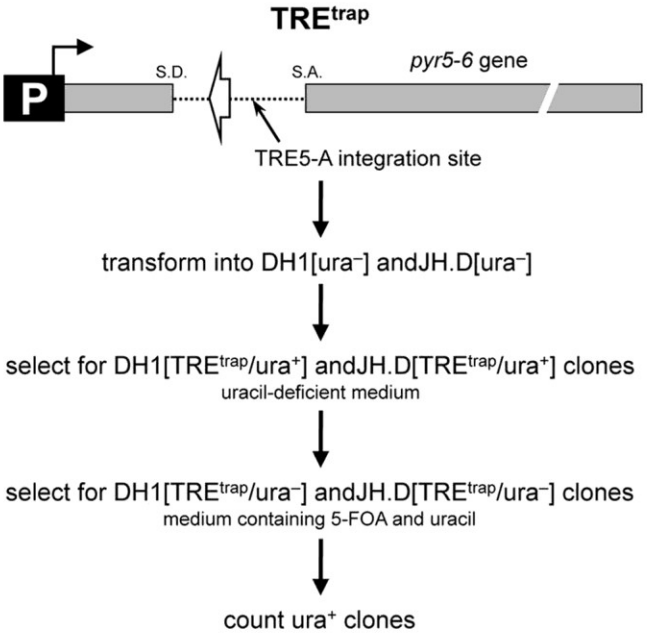


FIG. 4. Outline of the TRE trap assay.

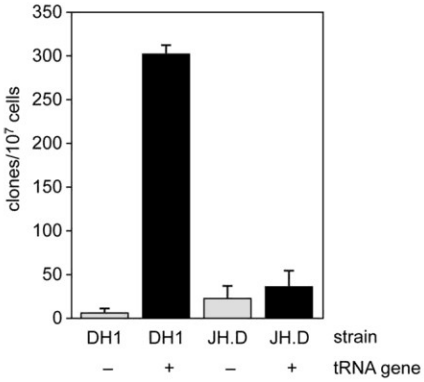


FIG. 5. TRE5-A retrotransposition activity in JH.D cells. TRE^{trap} plasmids without a tRNA gene as an integration target (gray bars) or with a *D. discoideum* Val^{UAC} tRNA gene as bait to attract integration of mobile TREs (black bars) were transformed into DH1[ura$^-$] or JH.D[ura$^-$] cells. Mean clone numbers (± SD) from 5 petri dishes are shown.

sensitive to the cytostatic drug 5-FOA and were killed unless they acquired mutations in the TRE^{trap} gene. Under selection with 5-FOA and uracil, resistant ura$^-$ clones appeared as a consequence of the disruption of the TRE^{trap} gene by the *de novo* integration of endogenous active TRE5-A elements (Fig. 5). As a negative control, we used a TRE^{trap} plasmid that lacked the bait tRNA gene. Under these conditions, the TRE trap bait did not attract TRE5-A retrotransposons; only a few clones were obtained (Fig. 5), and these probably arose from spontaneous mutations of the TRE^{trap} gene rather than targeted integration of TRE5-A elements (data not shown; discussed in reference 2).

We developed ura$^-$ mutants in CbfA-depleted JH.D cells by prolonged selection of cells in the presence of 5-FOA and uracil. For further experiments, we chose a JH.D[ura$^-$] mutant that had a complete chromosomal deletion of the *pyr5-6* coding region, as determined by PCR analysis (data not shown). We transformed TRE^{trap} plasmids into the JH.D[ura$^-$] cells and subjected JH.D[ura$^+$/TRE^{trap}] transformants to 5-FOA selection. As shown in Fig. 5, CbfA depletion of *D. discoideum* cells resulted in a >90% reduction in TRE5-A retrotransposition activity. We noticed a reproducible increase in background ura$^-$ clones obtained from JH.D[ura$^+$/TRE^{trap}] transformants after 5-FOA selection compared to those obtained from DH1[ura$^+$/TRE^{trap}] cells in the absence of a target tRNA gene (Fig. 5). This increase may reflect a generally higher mutation rate in CbfA-depleted cells, but this was not analyzed further.

Although these data were promising, they were challenged by the different strain histories of the cells used in this experiment: whereas the ura$^-$ mutant DH1 (the wild-type strain in this experiment) was derived from the axenic *D. discoideum* strain AX3, our CbfA mutant strain (JH.D) was derived from AX2. Although AX3 and AX2 have the same parent strain, NC4, the different laboratory histories of both strains may have corrupted TRE5-A retrotransposition activity in AX2 cells, thus mimicking a CbfA effect on TRE5-A retrotransposition in JH.D cells. We therefore decided to perform complementation studies in JH.D[ura$^-$] cells expressing plasmid-borne CbfA to rescue the aberrant TRE5-A retrotransposition activities. In two parallel series of transformations, TRE^{trap} plasmids that either contained or lacked a target tRNA gene were introduced into JH.D[ura$^-$] cells. The resulting transformants were supertransformed with plasmids that conferred resistance to G418 and supported the expression of CbfA$^{2\text{-}998}$ or CbfA-CTD. First, stable transformants were screened for comparable expression of the plasmid-borne CbfA variants (data not shown). Next, we measured the relative expression levels of TRE5-A in the transformants by quantitative RT-PCR (Fig. 6A). As expected from our data described above, we found that the expression of CbfA$^{2\text{-}998}$ or CbfA-CTD increased the expression of TRE5-A in cells carrying the TRE^{trap} gene. Finally, we measured the retrotransposition activity of the TRE5-A population in the TRE trap assay. The natural mutation rate observed for the TRE^{trap} gene was approximately 45 clones per 10^7 cells (Fig. 6B). This rate was not increased by the ectopic expression of CbfA$^{2\text{-}998}$ or CbfA-CTD (Fig. 6B). In contrast, when a target tRNA gene was included in the TRE^{trap} gene, the mutation rate at the TRE^{trap} gene increased approximately 4-fold in the presence of full-length CbfA, to roughly the wild-type levels observed in DH1 cells. This observation indicated full rescue of retrotransposition activity of the TRE5-A population. Unexpectedly, we found that expression of CbfA-CTD, although fully restoring aberrant TRE5-A expression, did not rescue the retrotransposition deficiency of the CbfA mutant. This result was observed in three independent experiments, and the reason for it remains obscure.

DISCUSSION

CbfA is a host factor that supports TRE5-A expression. In this report, we describe experiments demonstrating that CbfA is a host factor for TRE5-A expression in growing *D. discoideum* cells. In contrast to our prediction, we found that CbfA does not act as a component of the cellular defense machinery that is expected to limit the expression and subsequent ampli-

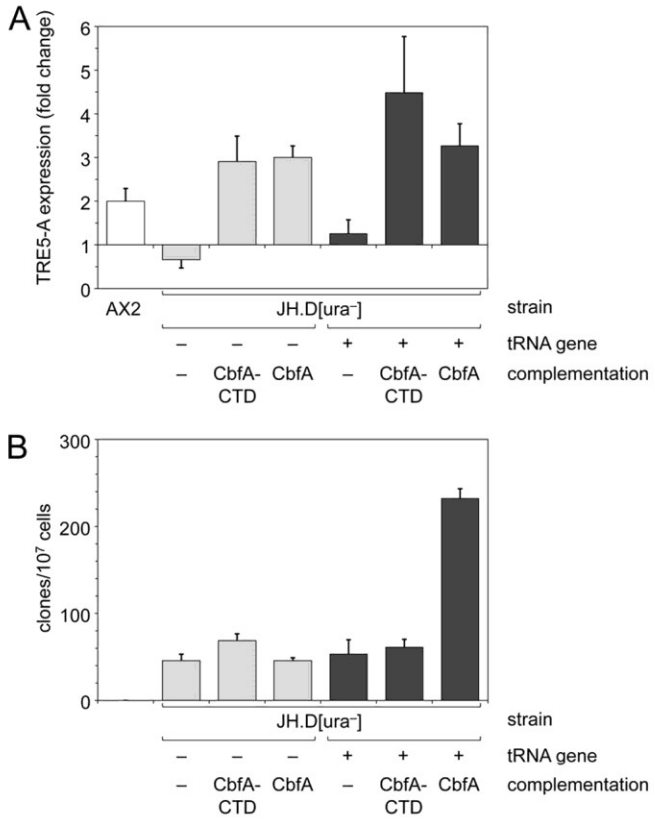


FIG. 6. Complementation of TRE5-A retrotransposition in the CbfA mutant. (A) TRE5-A.1 expression was quantified in JH.D[TRE^{trap}/ura^{+}] transformants by real-time RT-PCR. Expression of TRE5-A in AX2 cells versus untransformed JH.D[ura^{-}] cells served as a control (white bar, column 1). JH.D[ura^{-}] cells were transformed with TRE^{trap} plasmids without a tRNA gene as an integration target (gray bars, columns 2 to 4) or with a *D. discoideum* Val^{UAC} tRNA gene as bait to attract mobile TREs (black bars, columns 5 to 7). Resulting JH.D[TRE^{trap}/ura^{+}] cells were then supertransformed with empty expression vector (columns 2 and 5) or with plasmids that allowed for the expression of either CbfA-CTD (columns 3 and 6) or full-length CbfA (columns 4 and 7) in the JH.D background. All expression data were obtained from 4 to 8 independent transformants and are given relative to untransformed JH.D[ura^{-}] cells, meaning that values of >1 indicate more TRE5-A expression in the transformants or AX2 cells than in untransformed JH.D[ura^{-}] cells. (B) JH.D[TRE^{trap}/ura^{+}] cells were subjected to selection in 5-FOA and uracil. Numbers of 5-FOA-resistant clones were calculated for 5 petri dishes and are given as means ± SD. Columns are the same as in panel A. The experiment was repeated twice with similar results.

fication of retrotransposons in order to maintain genome integrity. CbfA has a strong supportive effect on the expression of both (+) RNA and (−) RNA of TRE5-A, but we found no pronounced differences in the promoter activity of the A module or the C module in CbfA-depleted cells, despite the sharp decrease in steady-state transcript levels of TRE5-A RNA in the mutant.

The data suggest that CbfA influences TRE5-A expression indirectly. An attractive hypothesis is that steady-state levels of TRE5-A RNA are regulated by posttranscriptional gene silencing (PTGS). Interestingly, we observed that the exposure of cells to antimycin A, and also the presence of a functional CbfA protein, stabilized both (+) RNA and (−) RNA of TRE5-A. In some organisms, for example, *Drosophila*, RNA interference (RNAi) is strictly ATP dependent (30), while in other organisms, such as humans (as demonstrated with human cells), RNAi seems to function in an ATP-independent manner (31). For *D. discoideum*, it is postulated that all known PTGS effects are mediated by RNA-dependent RNA polymerases (20), meaning that PTGS requires highly energy-consuming RNA-amplifying steps to work. Thus, assuming that antimycin A limits the energy metabolism in *D. discoideum* cells by blocking respiratory chain function, one could argue that antimycin affects TRE5-A transcript levels by blocking PTGS mechanisms. The role of CbfA in the stabilization of (+) RNA and (−) RNA of TRE5-A remains elusive. We speculate that CbfA regulates the expression of a gene involved in PTGS. This question will be addressed in future studies.

CbfA is essential for TRE5-A retrotransposition. Although CbfA-CTD is required and sufficient to maintain high TRE5-A transcript levels in *D. discoideum* cells, it was not able to restore the diminished retrotransposition of TRE5-A in the CbfA mutant. This result suggested a role for CbfA in the retrotransposition process in addition to the regulation of TRE5-A expression. Is CbfA a component of the TRE5-A preintegration complex? All available data indicate that TRE5-A is a canonical non-LTR retrotransposon that integrates via target-primed reverse transcription (TPRT), which means that no free double-stranded DNA intermediates of TRE5-A are supposed to occur in the cell to which CbfA may bind through its DNA-binding capacity. It also seems improbable that CbfA binds to TRE5-A-derived RNA because CbfA does not contain predictable RNA-binding domains or possess other enzymatic functions presumably required for the reverse transcription and integration processes. Although productive integration of mobilized retrotransposons may rely on the activity of host factors such as DNA repair enzymes, the reverse transcription and integration of such elements are likely to be mediated exclusively by TRE5-A ORF1/ORF2 functions. We concluded that it is unlikely that host-encoded CbfA has a direct function in the TPRT process.

Instead, the retrotransposition frequency of TRE5-A may be affected indirectly by CbfA if a component of the cellular DNA repair machinery exists that is expressed in a CbfA-dependent manner and cannot be controlled by CbfA-CTD alone. Support for this assumption may be deduced from the observation that we reproducibly obtained more ura^{-} mutants in the TRE trap assay with JH.D cells than in that with DH1 wild-type cells in the absence of a tRNA gene (columns 1 and 3 in Fig. 5), a sign of an increased general mutation rate in the mutant. However, we assume that nonproductive attempts by TRE5-A to integrate upstream of the bait tRNA gene in the TRE trap assay may have caused substantial genome instability at this locus due to insufficient DNA repair and, in turn, should have generated excess ura^{-} mutants. However, excess ura^{-} mutants were not observed for the CbfA mutant (column 4 in Fig. 5). On the contrary, we found that retrotransposon-induced mutagenesis of the TRE^{trap} gene was almost completely lost in the absence of CbfA.

One could imagine that the normal cellular function of CbfA as a putative chromatin-remodeling enzyme may provide a chromatin architecture in the vicinity of tRNA genes that sup-

ports the integration process. This is speculative as long as the putative functions of the JmjC and zinc finger domains of CbfA remain elusive.

In conclusion, CbfA is a host-encoded factor that strongly supports TRE5-A amplification in modern *D. discoideum* strains by controlling the absolute amount of transcript that is available for translation of TRE5-A proteins, reverse transcription of TRE5-A (+) RNA, and subsequent integration into chromosomal loci. CbfA has two distinct, indirect functions that can be assigned to different parts of the protein: maintenance of high steady-state levels of TRE5-A transcripts requires only the carboxy-terminal domain of CbfA, whereas the remaining parts of the protein, perhaps utilizing chromatin-remodeling activity of CbfA, are required to pass through a complete retrotransposition cycle.

ACKNOWLEDGMENTS

We are grateful to W. Nellen for providing the *rrpA* to *-C* mutants. This work was supported by the German Research Foundation DFG (WI 1142/6-1).

REFERENCES

1. **Agger, K., J. Christensen, P. A. Cloos, and K. Helin.** 2008. The emerging functions of histone demethylases. Curr. Opin. Genet. Dev. **18:**159–168.
2. **Beck, P., T. Dingermann, and T. Winckler.** 2002. Transfer RNA gene-targeted retrotransposition of *Dictyostelium* TRE5-A into a chromosomal UMP synthase gene trap. J. Mol. Biol. **318:**273–285.
3. **Caterina, M. J., J. L. S. Milne, and P. N. Devreotes.** 1994. Mutation of the third intracellular loop of the cAMP receptor, cAR1, of *Dictyostelium* yields mutants impaired in multiple signaling pathways. J. Biol. Chem. **269:**1523–1532.
4. **Chung, T., O. Siol, T. Dingermann, and T. Winckler.** 2007. Protein interactions involved in tRNA gene-specific integration of *Dictyostelium discoideum* non-long terminal repeat retrotransposon TRE5-a. Mol. Cell. Biol. **27:**8492–8501.
5. **Cost, G. J., Q. Feng, A. Jacquier, and J. D. Boeke.** 2002. Human L1 element target-primed reverse transcription in vitro. EMBO J. **21:**5899–5910.
6. **Craig, N. L., R. Craigie, M. Gellert, and A. M. Lambowitz (ed.).** 2002. Mobile DNA II. ASM Press, Washington, DC.
7. **Eichinger, L., et al.** 2005. The genome of the social amoeba *Dictyostelium discoideum*. Nature **435:**43–57.
8. **Fey, P., et al.** 2009. dictyBase—a *Dictyostelium* bioinformatics resource update. Nucleic Acids Res. **37:**D515–D519.
9. **Geier, A., J. Horn, T. Dingermann, and T. Winckler.** 1996. A nuclear protein factor binds specifically to the 3′-regulatory module of the long-interspersed-nuclear-element-like *Dictyostelium* repetitive element. Eur. J. Biochem. **241:**70–76.
10. **Horn, J., et al.** 1999. A *Dictyostelium* protein binds to distinct oligo(dA) · oligo(dT) DNA sequences in the C-module of the retrotransposable element DRE. Eur. J. Biochem. **265:**441–448.
11. **Klose, R. J., E. M. Kallin, and Y. Zhang.** 2006. JmjC-domain-containing proteins and histone demethylation. Nat. Rev. Genet. **7:**715–727.
12. **Ling, A. Z., R. B. Guyer, and R. A. Deering.** 2001. *Dictyostelium discoideum* plasmid containing an AP-endonuclease upstream sequence: bleomycin induction of a luciferase reporter. Environ. Mol. Mutagen. **38:**244–247.
13. **Luan, D. D., M. H. Korman, J. L. Jakubczak, and T. H. Eickbush.** 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell **72:**595–605.
14. **Lucas, J., et al.** 2009. The carboxy-terminal domain of *Dictyostelium* C-module-binding factor is an independent gene regulatory entity. PLoS One **4:**e5012.
15. **Malik, H. S., and T. H. Eickbush.** 1999. Modular evolution of the integrase domain in the Ty3/gypsy class of LTR retrotransposons. J. Virol. **73:**5186–5190.
16. **Malik, H. S., and T. H. Eickbush.** 2002. Origins and evolution of retrotransposons, p. 1111–1144. *In* N. L. Craig, M. Gellert, and A. M. Lambowitz (ed.), Mobile DNA II. ASM Press, Washington, DC.
17. **Maniak, M., and W. Nellen.** 1989. pISAR, a tool for cloning genomic sequences adjacent to the site of vector integration. Nucleic Acids Res. **17:**4894.
18. **Marschalek, R., J. Hofmann, G. Schumann, and T. Dingermann.** 1992. Two distinct subforms of the retrotransposable DRE element in NC4 strains of *Dictyostelium discoideum*. Nucleic Acids Res. **20:**6247–6252.
19. **Marschalek, R., J. Hofmann, G. Schumann, R. Gosseringer, and T. Dingermann.** 1992. Structure of DRE, a retrotransposable element which integrates with position specificity upstream of *Dictyostelium discoideum* tRNA genes. Mol. Cell. Biol. **12:**229–239.
20. **Martens, H., et al.** 2002. RNAi in *Dictyostelium*: the role of RNA-directed RNA polymerases and double-stranded RNase. Mol. Biol. Cell **13:**445–453.
21. **Pfaffl, M. W.** 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. **29:**e45.
22. **Schumann, G., I. Zündorf, J. Hofmann, R. Marschalek, and T. Dingermann.** 1994. Internally located and oppositely oriented polymerase II promoters direct convergent transcription of a LINE-like retroelement, the *Dictyostelium* repetitive element, from *Dictyostelium discoideum*. Mol. Cell. Biol. **14:**3074–3084.
23. **Siol, O., et al.** 2006. Role of RNA polymerase III transcription factors in the selection of integration sites by the *Dictyostelium* non-long terminal repeat retrotransposon TRE5-A. Mol. Cell. Biol. **26:**8242–8251.
24. **Siol, O., T. Dingermann, and T. Winckler.** 2006. The C-module DNA-binding factor mediates expression of the *Dictyostelium* aggregation-specific adenylyl cyclase ACA. Eukaryot. Cell **5:**658–664.
25. **Szafranski, K., et al.** 1999. Non-LTR retrotransposons with unique integration preferences downstream of *Dictyostelium discoideum* transfer RNA genes. Mol. Gen. Genet. **262:**772–780.
26. **Tsukada, Y., et al.** 2006. JmjC-domain-containing proteins and histone demethylation. Nature **439:**811–816.
27. **Winckler, T., et al.** 2004. CbfA, the C-module DNA-binding factor, plays an essential role in the initiation of *Dictyostelium discoideum* development. Eukaryot. Cell **3:**1349–1358.
28. **Winckler, T., K. Szafranski, and G. Glöckner.** 2005. Transfer RNA gene-targeted integration: an adaptation of retrotransposable elements to survive in the compact *Dictyostelium discoideum* genome. Cytogen. Genome Res. **110:**288–298.
29. **Winckler, T., et al.** 2001. Gene function analysis by amber stop codon suppression: CMBF is a nuclear protein that supports growth and development of *Dictyostelium* amoebae. J. Mol. Biol. **305:**703–714.
30. **Zamore, P. D., T. Tuschl, P. A. Sharp, and D. P. Bartel.** 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell **101:**25–33.
31. **Zhang, H., F. A. Kolb, V. Brondani, E. Billy, and W. Filipowicz.** 2002. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. EMBO J. **21:**5875–5885.

6.2 Manuskript 2

Conserved Gene Regulatory Function of the Carboxy-Terminal Domain of Dictyostelid C-Module-Binding Factor

---

Anika Schmith, Marco Groth, Josephine Ratka, Sara Gatz, Thomas Spaller, Oliver Siol, Gernot Glöckner, Thomas Winckler

Publiziert in: *Eukaryotic Cell*, 2013

EC
Journals.ASM.org

Conserved Gene Regulatory Function of the Carboxy-Terminal Domain of Dictyostelid C-Module-Binding Factor

Anika Schmith,[a] Marco Groth,[b] Josephine Ratka,[a] Sara Gatz,[a] Thomas Spaller,[a] Oliver Siol,[a,c] Gernot Glöckner,[d,e] Thomas Winckler[a]

Department of Pharmaceutical Biology, Institute of Pharmacy, University of Jena, Jena, Germany[a]; Genome Analysis, Leibniz Institute for Age Research—Fritz Lipmann Institute, Jena, Germany[b]; Institut de Génétique Humaine, CNRS, UPR 1142, Montpellier, France[c]; Institute for Biochemistry I, Medical Faculty, University of Cologne, Cologne, Germany[d]; Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany[e]

C-module-binding factor A (CbfA) is a jumonji-type transcription regulator that is important for maintaining the expression and mobility of the retrotransposable element TRE5-A in the social amoeba *Dictyostelium discoideum*. CbfA-deficient cells have lost TRE5-A retrotransposition, are impaired in the ability to feed on bacteria, and do not enter multicellular development because of a block in cell aggregation. In this study, we performed Illumina RNA-seq of growing CbfA mutant cells to obtain a list of CbfA-regulated genes. We demonstrate that the carboxy-terminal domain of CbfA alone is sufficient to mediate most CbfA-dependent gene expression. The carboxy-terminal domain of CbfA from the distantly related social amoeba *Polysphondylium pallidum* restored the expression of CbfA-dependent genes in the *D. discoideum* CbfA mutant, indicating a deep conservation in the gene regulatory function of this domain in the dictyostelid clade. The CbfA-like protein CbfB displays ~25% sequence identity with CbfA in the amino-terminal region, which contains a JmjC domain and two zinc finger regions and is thought to mediate chromatin-remodeling activity. In contrast to CbfA proteins, where the carboxy-terminal domains are strictly conserved in all dictyostelids, CbfB proteins have completely unrelated carboxy-terminal domains. Outside the dictyostelid clade, CbfA-like proteins with the CbfA-archetypical JmjC/zinc finger arrangement and individual carboxy-terminal domains are prominent in filamentous fungi but are not found in yeasts, plants, and metazoans. Our data suggest that two functional regions of the CbfA-like proteins evolved at different rates to allow the occurrence of species-specific adaptation processes during genome evolution.

Cells of the social amoebae (dictyostelids) live in the soil and feed on bacteria. Under unfavorable conditions, dictyostelid cells can aggregate and form multicellular fruiting bodies that hold dormant spores to ensure the survival of the species (1, 2). Insight into the genome structures of related dictyostelids has revealed an unexpected degree of genome flexibility and genetic diversity, suggesting that the study of these organisms may be valuable in understanding the evolutionary forces that drive genomic adaptation processes (3–5).

The model species *Dictyostelium discoideum* has an unusual genome, where only ~22% of the nucleotides are GC and 65% of the DNA codes for proteins (3). Compared with other dictyostelid genomes that present similar gene-dense environments (4), the *D. discoideum* genome harbors a surprisingly high percentage of mobile genetic elements (6). Transposon activity can be disastrous for the host if it leads to insertional mutagenesis, nonallelic homologous recombination, or induction of chromosome breaks (7, 8). Therefore, the activity of these parasitic elements must be carefully controlled by the host to maintain genome stability. *D. discoideum* has emerged as an excellent model to study the interactions between transposable elements and compact host genomes (9).

The retrotransposon TRE5-A is characterized by three outstanding features: an active population of elements in the *D. discoideum* genome is maintained (10), gene disruption rarely occurs because integration is targeted to the vicinity of tRNA genes (11), and a considerable amount of minus strand ("antisense") RNA is produced from a promoter located at the 3′ end of the element, the C module (12, 13). The large amount of minus strand RNA in growing *D. discoideum* cells suggests that TRE5-A retrotransposition may be regulated by posttranscriptional silencing. The C-module-binding factor A (CbfA) was discovered in a search for cellular factors that bind to the C module *in vitro* and may modulate TRE5-A amplification by regulating the TRE5-A minus strand of RNA (14).

Attempts to delete CbfA from *D. discoideum* cells by gene replacement have been unsuccessful. As an alternative, a *cbfA* gene knock-in mutant was generated in which the codon at position 455 was replaced with a TAG (*amber*) stop codon (15). Because of the low frequency of *amber* codons in the *D. discoideum* genome, the expression of an *amber* suppressor tRNA gene does not produce a phenotype (16) but rather allows readthrough translation of the full-length protein from the *cbfA* (*amber*) mRNA. Because of the low efficacy of *amber* suppression, the JH.D strain produces less than 5% of the full-length CbfA protein (15).

In cells of the CbfA mutant JH.D, both plus- and minus-strand RNA from the retrotransposon TRE5-A is diminished and the mobility of the endogenous retrotransposon population is drastically reduced (13). Thus, CbfA is a positive regulator of TRE5-A amplification. Expression and retrotransposition of TRE5-A can be restored in JH.D cells by the ectopic expression of the full-length CbfA protein. Interestingly, the expression of the isolated carboxy-terminal domain (CTD) of CbfA in the mutant cells was sufficient for complete restoration of the TRE5-A transcript levels, even though TRE5-A retrotransposition was unaffected (13).

Received 23 November 2012 **Accepted** 18 January 2013
Published ahead of print 25 January 2013
Address correspondence to Thomas Winckler, t.winckler@uni-jena.de.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/EC.00329-12.

doi:10.1128/EC.00329-12

JH.D cells are defective in phagocytosis, cytokinesis (15), and multicellular development (17). In mutant cells, the expression of the aggregation-specific adenylyl cyclase ACA is strongly reduced; therefore, cyclic-AMP (cAMP)-induced gene expression is absent. However, multicellular development of JH.D cells can be restored either by the application of exogenous cAMP pulses or by mixing mutant and wild-type cells, suggesting that the developmental phenotype of JH.D cells is not cell autonomous (17). The aggregation block in JH.D cells can be overcome by the ectopic expression of the full-length CbfA protein (17), but not by the expression of the isolated CbfA CTD (18).

The observation that the CTD of CbfA restored TRE5-A expression raised questions concerning the origin of this domain and its role as an independent gene regulatory entity in CbfA function. In this study, we show that CbfA is a general transcriptional regulator in growing *D. discoideum* cells and that most CbfA-dependent gene expression is mediated by the CbfA CTD. The CbfA protein, including the CTD, is highly conserved throughout dictyostelid evolution. Dictyostelids contain a CbfB protein. This protein is paralogous to CbfA in the amino-terminal portion of the protein but has an individual CTD with no similarity to those of other CbfA or CbfB proteins. We speculate that the two gene regulatory elements within the CbfB proteins in individual species have evolved differently to facilitate the emergence of genome-specific gene regulatory functions.

MATERIALS AND METHODS

Dictyostelids. The dictyostelid species used in this study were obtained from the *Dictyostelium* Stock Center, which can be accessed at dictyBase (http://dictybase.org). The following strains were used (dictyBase strain identification numbers are in parentheses): *D. citrinum* OH594 (DBS0235738), *D. rosarium* M45 (DBS0235885), *D. dimigraformum* AR5B (DBS0235745), *D. giganteum* WS589 (DBS0235820), *D. sphaerocephalum* GR11 (DBS0235889), *D. robustum* Ch53 (DBS0235883), *D. firmibasis* TNS-C-14 (DBS0235812), *D. intermedium* PJ11 (DBS0235829), *D. longosporum* TNS-C-109 (DBS0235836), *D. clavatum* TNS-C-189 (DBS0235739), *D. brunneum* WS700 holotype (DBS0235734), *D. brefeldianum* G-12-1 (DBS0235732), *D. minutum* 71 (DBS0235843), *D. lacteum* (DBS0235831), *Polysphondylium pallidum* WS320 (ATCC 44843), *Acytostelium subglobosum* Lb1 (DBS0235452), and *D. fasciculatum* SH3 (DBS0235810).

Isolation of the dictyostelid *cbfA* and *cbfB* gene sequences. The dictyostelid species were grown on SM/5 agar plates (19) in association with *Raoultella planticola*. Before the first multicellular stages appeared, the cells were harvested, washed, and stored as pellets consisting of 2×10^7 cells. The pellets were frozen at −80°C. Genomic DNA was prepared from the frozen cells with the Qiagen DNeasy Tissue kit. Total RNA was prepared from the frozen cells with the Qiagen RNeasy Minikit. cDNA synthesis was performed with the Qiagen Omniscript RT kit with an oligonucleotide-deoxythymidine primer.

Degenerate primers were derived from the *D. discoideum* CbfA protein sequence and multiple sequence alignments of CbfA sequences from group 4 species. DNA sequences of the orthologous *cbfA* genes were generated from overlapping PCR fragments. The PCR products were either cloned into the pGEM-T plasmid (Promega) or sequenced directly from the PCR fragments. The DNA sequences representing approximately 90% of the encoded CbfA proteins were assembled and aligned with COBALT (20).

The *D. discoideum cbfA* and *cbfB* gene sequences can be accessed via dictyBase (http://dictybase.org/) under accession numbers DDB_G0279409 and DDB_G0293470, respectively.

Information about the *cbfA* and *cbfB* genes of *D. fasciculatum* and *P. pallidum* was assembled in the Social Amoebas Comparative Genome Browser (http://sacgb.fli-leibniz.de/cgi/index.pl) from the preliminary data of the comparative genome sequencing project (4). CbfA and CbfB sequences of *A. subglobosum* were generously provided by the *Acytostelium* gene analysis team at the University of Tsukuba (H. Urushihara, personal communication). The *cbfA* and *cbfB* sequences from *D. purpureum* were generously provided by A. Kuspa (Baylor College of Medicine, Houston, TX). These sequences are now publicly available at dictyBase (5). The structures of the *cbfA* and *cbfB* gene sequences as predicted by the dictyostelid comparative genome projects were confirmed by sequencing overlapping PCR products that were generated from both genomic DNA and cDNA of the respective species.

Expression of the CbfA CTD mutant proteins. A gene fragment spanning amino acids 724 to 1000 of the *D. discoideum* CbfA protein was cloned as a BamHI fragment into the pDXA-GFP2 plasmid (21). The resulting plasmid, pDXA-DdCbfA-CTD, was later transformed into JH.D cells (15). The corresponding sequence of *P. pallidum* CbfA was isolated by PCR and cloned as mentioned above. All of the strains were cultured in standard HL-5 medium (ForMedium, Norfolk, United Kingdom) supplemented with 10 μg/ml G418. Live cells expressing the green fluorescent protein (GFP)-tagged CbfA CTD were stained with 20 μg/ml 4′,6-diamidino-2-phenylindole (DAPI) in phosphate-buffered saline (PBS) for 30 min at room temperature. The cells were washed with PBS and observed with an Axio Observer Z1 microscope (Carl Zeiss AG). The expression of GFP-tagged CbfA CTD proteins was analyzed by Western blotting as described previously (22).

Illumina RNA sequencing (RNA-seq). Total RNA was prepared from cells grown in three independent cultures containing AX2 cells (wild type), JH.D cells (the CbfA-underexpressing mutant), or JH.D cells expressing the GFP-tagged CbfA CTD from *D. discoideum* or *P. pallidum*. For library preparation, 5 μg of total RNA per sample was processed with an Illumina mRNA-Seq sample prep kit (RS-122-2001) by following the manufacturer's instructions. Library preparation was performed with the total RNA from the three independent cultures. The libraries were sequenced with a HiSeq 2000 instrument in a single-read approach with 50 cycles resulting in reads with a length of 50 nucleotides. Reads were mapped to the *D. discoideum* reference genome downloaded from dictyBase (http://dictybase.org/) with the Illumina-supplied tool ELAND (23) by using standard parameters. The expression values were calculated by determining the number of uniquely mappable reads per transcript and normalizing that value to the total number of mappable reads and the length of the respective transcript (adapted from reference 24). This approach resulted in RPKM values (reads per kilobase transcript and million mappable reads) representing the expression level of the respective transcript. The determination of differentially expressed genes was performed with the unnormalized count data with the R statistical computing (25) tools edgeR (26) and DEseq (27). Genes were considered to be differentially expressed if the *P* value was ≤0.01. Furthermore, an at least 3-fold difference in gene expression was considered CbfA-dependent regulation. Gene ontology (GO) term enrichment analyses were performed with AmiGO version 1.8 (28).

Phylogenetic analysis. The multiple-sequence alignment was performed with ClustalX 2.0 (29). The phylogenetic tree was generated with Tree-Puzzle 5.2 (30) under the Dayhoff model. The heterogeneity rate was uniform. Supporting values are derived from 1,000 puzzling steps. The tree was rooted with the *Naegleria* sequence as the outgroup.

Quantitative reverse transcription (qRT)-PCR. Total RNA was prepared from frozen pellets consisting of 2×10^7 cells with the Qiagen RNeasy Minikit according to the protocol provided. cDNA was synthesized by the RT of 500 ng of total RNA with an oligo(dT) primer and the Qiagen Omniscript RT kit. Real-time amplification of the genes was performed with the *Taq* polymerase mix from Jena Bioscience (Jena, Germany) supplemented with EvaGreen Fluorescent Gel Stain and ROX Reference Dye (both from Jena Bioscience, Jena, Germany) on a Stratagene Mx3000P instrument. The sequences of the primers used are listed in Table S1 in the supplemental material. All of the measurements were performed in triplicate. Real-time PCR signals were normalized to the

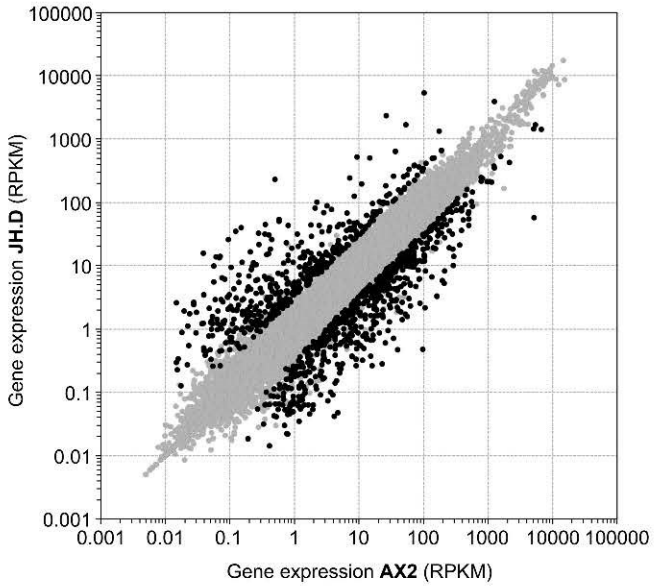


FIG 1 Transcriptome analysis of CbfA mutant JH.D cells. Shown are the results from an RNA-seq experiment with AX2 and JH.D cells. The reads were mapped to 12,317 gene models, and read counts were standardized to transcript lengths and 1 million reads (RPKM values). The mean RPKM values of three independent cultures are plotted as gray symbols. The black symbols highlight genes that were reported to show statistically significant differential levels of regulation ($P \leq 0.01$) of at least 3-fold.

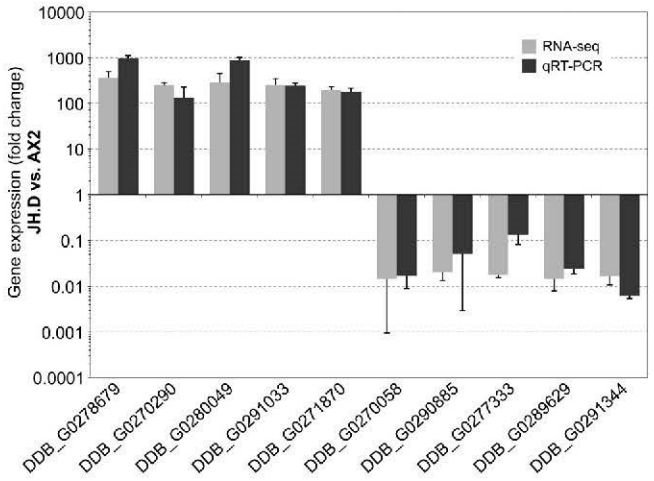


FIG 2 qRT-PCR analysis of a representative set of CbfA-regulated genes. Comparison of RNA-seq data (gray bars) with qRT-PCR data (black bars) for five CbfA-repressed and five CbfA-supported genes. The qRT-PCR was performed with cDNA prepared from three independent cultures of AX2 and JH.D cells that were different from the three cultures used for RNA-seq. Values are presented as means of three biological replicates ± the standard deviations. Note that the expression of genes is calculated for JH.D versus AX2 cells, meaning that values of >1 represent genes usually repressed by CbfA (and therefore overexpressed in JH.D), whereas values of <1 represent genes whose expression is supported by CbfA.

expression of the gene encoding *catA* (dictyBase entry DDB_G0274595). After an initial denaturing step at 95°C for 10 min, PCR was performed for 40 cycles at 95°C for 30 s, 60°C for 45 s and 72°C for 30 s. Differential gene expression was calculated according to the Pfaffl method (31) with *catA* as a reference gene.

Nucleotide sequence and RNA-seq data accession numbers. The partial *cbfA* gene sequences that were determined in this study have been deposited in GenBank under accession numbers JQ918362 (*D. giganteum cbfA*), JQ918358 (*D. citrinum cbfA*), JQ918363 (*D. intermedium cbfA*), JQ918366 (*D. sphaerocephalum cbfA*), JQ918359 (*D. clavatum cbfA*), JQ918360 (*D. dimigraformum cbfA*), JQ918361 (*D. firmibasis cbfA*), JQ918365 (*D. rosarium cbfA*), JQ918357 (*D. brunneum cbfA*), and JQ918356 (*D. brefeldianum cbfA*). The RNA-seq data have been deposited in the Gene Expression Omnibus database under accession number GSE38418.

RESULTS

CbfA is a general transcriptional regulator. The complex phenotype of CbfA-deficient mutants suggested that CbfA functions as a general transcriptional regulator in growing and developing *D. discoideum* cells. To test this hypothesis, we defined a list of CbfA-dependent genes in growing cells. To this end, we performed Illumina RNA-seq analyses on the *D. discoideum* wild-type strain AX2 and the CbfA-deficient mutant JH.D. RNA-seq on three independent cultures from each strain was highly reproducible (see Fig. S1 in the supplemental material) and yielded approximately 20 million sequence reads per RNA preparation, which were mapped to 12,317 *D. discoideum* gene models. We plotted the normalized sequence reads (RPKM values) for AX2 versus JH.D. Considerable differences in gene expression between the strains existed (Fig. 1). Statistical analysis revealed that 2,298 genes were differentially expressed between the AX2 and JH.D samples. As expected, the numbers of differentially expressed genes varied considerably, depending on the threshold setting (see Fig. S2 in the supplemental material). A 3-fold change in gene expression was considered to be indicative of CbfA-regulated gene expression. With this threshold, 1,030 genes were differentially regulated (see Table S2 in the supplemental material). Of these genes, 584 had reduced expression in JH.D cells, suggesting that these genes are positively regulated by CbfA. Fifty-seven genes were downregulated more than 20-fold in the mutant. The most pronounced CbfA-regulated gene was *abpF* (205-fold differential expression), which codes for an actin-binding protein that is most likely localized to the centrosomes. We found that 446 genes were upregulated in JH.D cells, which suggests that CbfA normally represses these genes. Eighty-six genes were upregulated by at least 20-fold in the CbfA mutant cells versus the AX2 cells. The most significant differences were a 458-fold change in the expression of a gene with unknown function (DDB_G0283511) and a 388-fold upregulation of *cyp513F1*, which encodes a cytochrome P450 enzyme. The gene *agnC*, which codes for an Argonaute protein that may be involved in processes involving posttranscriptional gene silencing, was upregulated by 225-fold in the CbfA mutant cells. We performed a qRT-PCR analysis of a representative set of 10 CbfA-regulated genes to verify the RNA-seq data. We measured the gene expression in the RNA preparations from 3 cultures that were grown independently from the cells used for RNA-seq. As indicated in Fig. 2, the qRT-PCR data confirmed the RNA-seq results.

GO term enrichment analysis with the program AmiGO set at a $P < 0.05$ cutoff indicated that a significant enrichment of the positively CbfA-regulated genes was involved in the assembly and maintenance of the actomyosin cytoskeleton and phagocytic vesicles, as well as general cellular processes, such as cytokinesis and cell morphogenesis (see Table S3 in the supplemental material). This observation correlated well with the growth characteristics of CbfA-deficient cells, including a decreased phagocytosis rate, slow growth on bacteria, and the generation of giant cells containing

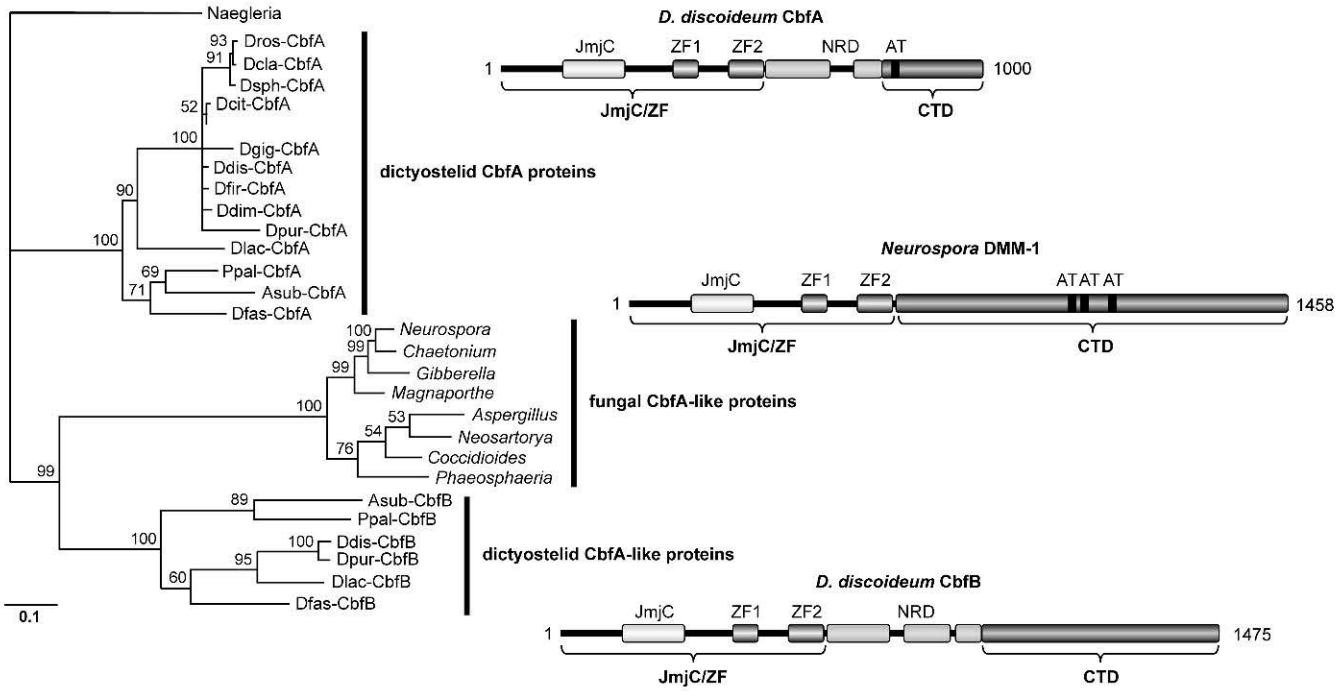


FIG 3 Phylogenetic relationships of CbfA-like proteins. Protein sequences corresponding to amino acids 1 to 553 of *D. discoideum* CbfA (JmjC/ZF region) were aligned with ClustalX 2.0 and analyzed with Tree-Puzzle 5.2. The values above the lines indicate the support values derived from 1,000 puzzling steps. In addition to the dictyostelid CbfA and CbfB proteins, the following fungal CbfA-like proteins were used in this alignment: *Aspergillus nidulans* (GI: 67528200), *Chaetonium globosum* (GI: 116199707), *Coccidioides immitis* (GI: 119188365), *Gibberella zeae* (GI: 46138433), *Magnaporthe grisea* (GI: 39973571), *Neosartorya fischeri* (GI: 119501326), *Neurospora crassa* (GI: 85078924), and *Phaeosphaeria nodorum* (GI: 111066574). Note that the *N. crassa* protein is identical to DMM-1, as reported by Honda et al. (39). The ZF2 region was also aligned with the zinc finger domain of *Arabidopsis* monoamine oxidase A repressor R1 (GI: 145360248) and a related zinc finger region in human cell division cycle-associated 7-like protein isoform 2 (GI: 188497635). Domain structures are schematically presented for *D. discoideum* CbfA, *D. discoideum* CbfB, and *N. crassa* DMM-1. JmjC, jumonji C domain; ZF, zinc finger region, AT, AT hook.

multiple nuclei (15). We performed a qRT-PCR analysis of 10 of these genes on RNA from three independently grown cultures to confirm the RNA-seq data. The genes selected exhibited between 3.3- and 31.2-fold higher expression levels in wild-type cells than in the CbfA mutant cells in the RNA-seq experiment. Surprisingly, the qRT-PCR data did not confirm the RNA-seq data (data not shown). This discrepancy, however, was not due to a general incongruence between RNA-seq and qRT-PCR data because the values for the genes measured by qRT-PCR in the RNA samples that we used for RNA-seq were comparable to the values measured by RNA-seq (not shown). Thus, it is probable that the expression of the genes putatively involved in the CbfA-dependent regulation of phagocytosis is generally variable and subject to subtle changes in culture conditions and the involvement of CbfA in this gene regulation remains uncertain. We should note, however, that the slow-growth phenotype of JH.D cells on bacteria lawns is reversed by the expression of CbfA in the mutant (data not shown).

CbfA is a highly conserved protein. *D. discoideum* CbfA is a 1,000-amino-acid multidomain protein (Fig. 3) that belongs to the family of jumonji-type transcriptional regulators. These proteins are often involved in deciphering the histone code by removing methyl groups from methyllysine or methylarginine residues in the histone tails (32, 33). A "carboxy-terminal jumonji domain" (JmjC) catalyzes this type of oxidative demethylation (reviewed in references 34 and 35). The JmjC domain is located in the amino-terminal portion of the CbfA protein, between amino acids 113 and 280. This domain is followed by two cysteine-rich, zinc finger-like regions located at positions 373 to 414 (ZF1) and 492 to 550 (ZF2). Downstream of the ZF motifs is a distinct region of 217 amino acids that consists of 50% asparagine residues; therefore, this region is referred to as the CbfA asparagine-rich domain (NRD). The CbfA NRD separates the JmjC/ZF region of the CbfA protein from the CTD (CbfA CTD), which spans 230 amino acids and contains a single AT hook.

Dictyostelids are classified into four major groups based on their phylogenetic relationships (36). *D. discoideum* belongs to the latest diverged group, 4. On the basis of the primary structure of *D. discoideum* CbfA (14), we obtained partial sequences of the *cbfA* genes from eight other group 4 species by PCR with degenerate primers. These data were complemented with full-length *cbfA* genes obtained from the recently completed genome sequencing projects for the group 4 species *D. purpureum* (5), the group 3 species *D. lacteum* (G. Glöckner, unpublished data), the group 2 species *P. pallidum* (4) and *A. subglobosum* (available in the *Acytostelium* Genome Database), and the group 1 member *D. fasciculatum* (4). Sequences of the *cbfA* genes derived from genome sequencing projects were confirmed by sequencing genomic DNA and cDNA (data not shown). An alignment of 14 dictyostelid CbfA proteins is presented in Fig. S3 in the supplemental material.

Within group 4, the level of identity between the JmjC/ZF domains ranges from 74 to 97% and covers ~550 amino acids. The CbfA proteins from the organisms in groups 1 to 3 are more than 45% identical to the JmjC/ZF region of CbfA of *D. discoideum*. CbfA-like zinc finger regions (see Fig. S4 in the supplemental ma-

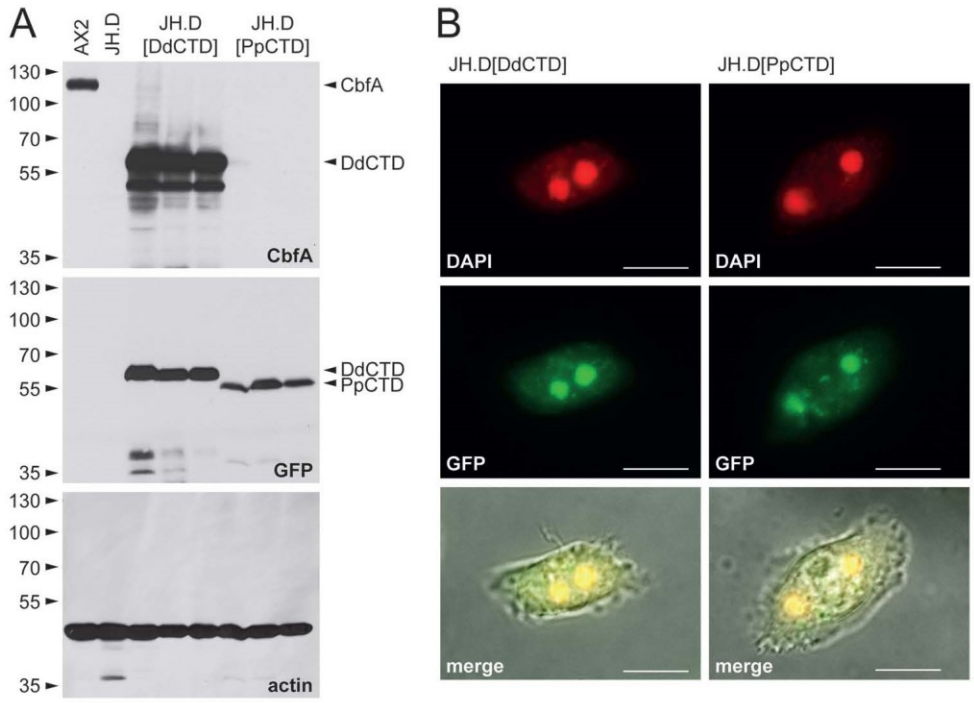


FIG 4 Expression of CbfA CTD in JH.D cells. (A) Western blot assays illustrating expression levels of the GFP-tagged *D. discoideum* CbfA CTD and the *P. pallidum* CbfA CTD in JH.D cells. Total cell extracts were prepared from AX2 and JH.D cells and the three JH.D transformants that were used for RNA-seq experiments. Shown is a single blot that was successively stained with CbfA monoclonal antibody 7F3 (which recognized *D. discoideum* CbfA CTD but not *P. pallidum* CbfA CTD), a GFP-specific antibody, and a polyclonal antiserum specific for actin (loading control). Values to the left indicate molecular mass markers (kilodaltons). (B) Cellular localization of the GFP-tagged *D. discoideum* CbfA CTD and the *P. pallidum* CbfA CTD in JH.D live cells. Top panel, DAPI staining; middle panel, GFP fluorescence; bottom panel, merged DAPI and GFP fluorescence overlaid with a phase-contrast picture. Note that a *D. discoideum* culture usually contains a fraction of multinucleated cells. Cells with two nuclei are compared here. Bars, 10 μm.

terial) are found in proteins from plants and animals; however, these proteins contain either ZF1 or ZF2 and are not linked with a JmjC domain. The CbfA-like combination of JmjC and ZF1/ZF2 domains was detected in proteins from several filamentous fungi (but not yeasts). Furthermore, this structure was detected in a protein of unknown function from the amoeboflagellate *Naegleria gruberi*. A protein with similarity to CbfA was also detected in the genome of the amoebozoan *Acanthamoeba castellanii* (Glöckner, unpublished). As discussed below, the CbfA-archetypical JmjC/ZF1/ZF2 design combined with the divergence of the CTDs defines a new family of "C-module-factor-like proteins."

CbfA CTDs are well conserved, with sequence identities ranging from 81 to 98% in group 4 and >40% in the non-group 4 species compared with *D. discoideum* CbfA. This *D. discoideum* CbfA CTD is later referred to as the CTD "core" because CbfA CTDs from non-group 4 species are considerably extended (see Fig. S3 in the supplemental material). CbfA CTD extensions range from 140, 223, 107, and 255 amino acids in CbfA from *D. lacteum*, *P. pallidum*, *A. subglobosum*, and *D. fasciculatum*, respectively. These extensions do not display any detectable sequence homology with each other, despite the strong conservation of the CTD core.

The CTD of CbfA displays deeply conserved gene regulatory functions. Having established a comprehensive list of CbfA-regulated genes in growing *D. discoideum* cells from the RNA-seq data, we next investigated whether the CbfA CTD could repair the aberrant gene expression in JH.D cells. The cells were transformed with a plasmid containing a GFP-tagged *D. discoideum* CbfA CTD variant under the control of a constitutive promoter. As shown in Fig. 4A, the expression of the CbfA CTD was higher than the endogenous CbfA levels in AX2 cells, and the CbfA CTD protein was enriched in the nuclei of JH.D transformants (Fig. 4B). These transformants were used for RNA-seq experiments, and gene expression levels were compared with those of AX2 and JH.D cells. Gene expression levels in JH.D cells that were 50% of the levels in AX2 cells were considered to be successfully restored by CbfA CTD transformation. Thus, 85% of the CbfA-activated and 76.7% of the CbfA genes were responsive to CbfA CTD. With a more stringent threshold of 80% of the wild-type expression level, we still determined that 76.5% of the CbfA-activated genes and 47.3% of the CbfA-repressed genes were complemented by the expression of the CbfA CTD in the CbfA mutant. We observed a pronounced overcomplementation in the mutant cell transformants; 69.5% of the CbfA-activated genes in the CbfA CTD-transformed mutant cells were expressed at levels higher than those observed in wild-type cells (see Table S2 in the supplemental material). The overcomplementation phenomenon in the mutant cells could result from the overexpression of the CbfA CTD. Alternatively, the JmjC/ZF region of CbfA may modulate the activity of the CTD in the full-length protein. In summary, the CbfA CTD may function to regulate the majority of the CbfA-regulated genes without requiring the JmjC/ZF domains.

We next determined whether the CbfA CTD of a dictyostelid species that diverged early would be functionally equivalent to the *D. discoideum* CbfA CTD. We cloned the CbfA CTD from *P. pallidum* and expressed the gene as a GFP-tagged protein in *D. dis-*

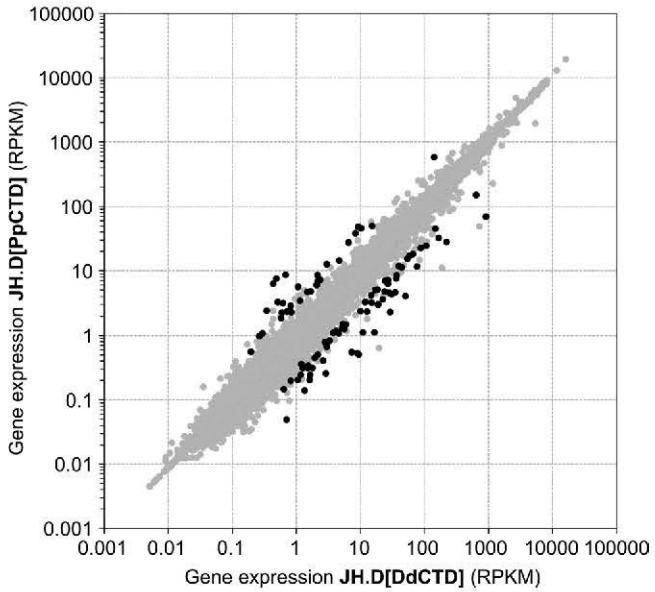


FIG 5 Gene regulatory function of *P. pallidum* CbfA CTD. Shown are the results from an RNA-seq experiment with JH.D cells expressing either the GFP-tagged *D. discoideum* CbfA CTD or the *P. pallidum* CbfA CTD. The reads were mapped to 12,317 gene models, and read counts were standardized to transcript lengths and 1 million reads (RPKM values). The mean RPKM values of three independent cultures are plotted (gray symbols). The black symbols highlight genes that were reported to show statistically significant differential levels of regulation ($P \leq 0.01$) of at least 3-fold.

coideum JH.D cells (Fig. 4A). Similar to the *D. discoideum* CbfA CTD, the *P. pallidum* CbfA CTD was localized to the nuclei of the transformed cells (Fig. 4B). RNA-seq data revealed that the expression of 73.3% of the CbfA-activated genes (73.3%) and 51.4% of the CbfA-repressed genes was restored by the *P. pallidum* CbfA CTD to at least 80% of the expression level in AX2 cells. Although the expression of the *P. pallidum* CbfA CTD in JH.D cells was considerably less than the expression of the *D. discoideum* CbfA CTD (Fig. 4A), we observed a similar overcomplementation effect on CbfA-dependent gene expression (64.0% of the CbfA-activated genes were expressed at >100% of the AX2 levels; see Table S2 in the supplemental material). Thus, overcomplementation by the CbfA CTD may, in fact, be a consequence of the missing JmjC/ZF region of the CbfA protein rather than an artifact of the overexpression process.

In Fig. 5, we plotted the RNA-seq data obtained from JH.D cells expressing the *D. discoideum* CbfA CTD versus that from the mutant cells expressing the *P. pallidum* CbfA CTD. Gene expression differences that were less than 3-fold were not considered for analysis. We determined that only 98 (0.8%) of the 12,317 genes were differentially regulated in the two data sets. Of these genes, only 31 were identified on the list of 1,030 CbfA-regulated genes, indicating a great functional overlap between the *D. discoideum* and *P. pallidum* CbfA CTDs. In fact, we found that approximately 87% of the CbfA CTD-responsive genes were restored to at least 80% of the wild-type level by both the *D. discoideum* and *P. pallidum* CbfA CTDs. This result is notable because *D. discoideum* and *P. pallidum* have different percentages of GC content in their genomes and promoter alignments of the two species are nearly impossible (4); however, the gene regulatory function of the *D. discoideum* CbfA CTD requires direct binding to the AT-rich DNA sequences within the target promoters, as suggested by *in vitro* and *in vivo* data (14, 18).

CbfA-like proteins contain conserved JmjC/ZF domains and divergent carboxy termini. "CbfA-like proteins," which we define in this study as proteins displaying the characteristic JmjC/ZF1/ZF2 architecture of dictyostelid CbfA, were detected both within and outside the dictyostelid clade. The average percentage of amino acid sequence identity between CbfA and CbfA-like proteins from dictyostelids (i.e., paralogous CbfB proteins) and fungi is approximately 25% in the protein regions with conserved architecture. The distinctive design of the zinc finger regions of CbfA (see Fig. S4 in the supplemental material) provides an indication of the evolution of CbfA and CbfA-like proteins. The ZF1 region of CbfB proteins shares the signature $CX_2CX_7\mathbf{X}X_2CX_2C$ (where **X** is any amino acid replacing cysteine in this position) with both fungal CbfA-like proteins and one of the earliest diverged dictyostelid, *D. fasciculatum*. This motif developed into $CX_2CX_7CX_2CX_2C$ in dictyostelid CbfA proteins and was further expanded by a CX_2H motif found exclusively in CbfA proteins (see Fig. S4 in the supplemental material). Importantly, a CbfA-like protein from *Naegleria gruberi* and CbfB from the dictyostelid *A. subglobosum* (group 2 dictyostelid) share a $CX_2CX_7CX_2\mathbf{X}X_2\mathbf{X}$ variant of ZF1 (**X** replacing cysteine in this position). In contrast to the design of the ZF1 region, that of the ZF2 region is highly conserved among CbfA, CbfB, and fungal CbfA-like proteins, although the distances between the two cysteine residues in the second CX_2C motif are extended by up to 36 amino acids in the CbfB proteins.

The arrangement of the cysteines in the ZF1 region suggests that the fungal CbfA-like proteins are more closely related to the dictyostelid CbfB proteins than to the CbfA proteins. A phylogenetic analysis of the JmjC/ZF regions of the CbfA-like proteins revealed a clear separation between the dictyostelid CbfA and CbfB proteins and their fungal orthologs (Fig. 3). To date, the CbfA CTD has not been found in proteins from any other organisms outside the dictyostelids. In striking contrast to the conservation of the CTDs of the CbfA proteins, the CbfB proteins have extended CTDs that do not exhibit any signs of similarity. The CbfB CTDs from the group 4 organisms *D. discoideum* and *D. purpureum* do not show any similarity, even though they are 85% identical in the JmjC/ZF regions. Similarly, the CTDs of the CbfA-like proteins from fungi do not share any traceable similarities. These observations indicate that the CTDs of the CbfA-like proteins have rapidly evolved because of either species-specific adaptation pressures or the total removal of clade-specific constraints.

DISCUSSION

Two different gene regulatory functions combined in one protein. In this study, we performed Illumina RNA-seq to extend our knowledge concerning the function of the CbfA protein. A previous DNA microarray experiment performed with CbfA-deficient mutant JH.D cells (18) had limited sensitivity and included only approximately half of the *D. discoideum* genes. Comparing the two transcriptomic approaches, we observed that only 35% of the 1,030 CbfA-dependent genes identified by RNA-seq were present on the DNA microarray. Only 42 of the 365 genes were reported as CbfA regulated in the DNA microarray experiment, where the threshold was set at a 1.5-fold change in expression, thereby suggesting that the sensitivity of the DNA microarray was lower than that of the RNA-seq approach. This assumption is supported by the observation that the detection of CbfA-dependent genes by DNA microarray analysis was biased toward strongly expressed

genes, as illustrated in Fig. S5 in the supplemental material. Both the DNA microarray experiments and the RNA-seq analyses reported CbfA positively regulated genes that were involved in phagocytosis. However, the results obtained with the two transcriptomic methods were not similar. GO term enrichment analysis of the DNA microarray data with AmiGO version 1.8, which is the same program used to analyze the RNA-seq data, reported that 16 genes that were assigned to the cellular function of "phagocytic vesicles" were regulated by CbfA (data not shown). However, nine of these genes produced conflicting results in the RNA-seq experiment and were not detected as CbfA-dependent genes. Even more confusing was the observation that qRT-PCR analysis measuring the expression of the CbfA-regulated genes that were associated with phagocytosis did not confirm the prediction from RNA-seq. Because we used different cultures for the RNA-seq and qRT-PCR experiments, we hypothesize that the expression of genes involved in phagocytosis is considerably variable and culture dependent.

Consistent with the previous DNA microarray data (18), we found with RNA-seq that the CbfA CTD core exerts the majority of its gene regulatory activity without requiring the other domains of CbfA. Thus, CbfA can be divided into two independently functioning regions: the JmjC/ZF region and the CTD. At present, we are not able to determine how the JmjC/ZF region of the CbfA protein contributes to CbfA function. In the experiments presented here, approximately two-thirds of the CbfA-mediated gene regulation was mediated by the CbfA CTD alone. Thus, the JmjC/ZF region of CbfA may only function as a regulatory domain to modulate the CbfA CTD. However, there is evidence that the JmjC/ZF portion of CbfA is required for gene regulatory functions that cannot be carried out by the CbfA CTD alone. For example, the CbfA CTD was not able to rescue the developmental defect in CbfA-deficient JH.D cells (18). The fact that CbfA-deficient cells have a strong developmental phenotype indicates that CbfB, which is expressed in the CbfA mutant at normal levels, is not able to complement the loss of CbfA. Nevertheless, we cannot exclude the possibility that the number of genes regulated by the JmjC/ZF region of CbfA has been underestimated in our study because of a partial overlap of CbfA with similar functions of the CbfB-derived JmjC/ZF domain. This question will be addressed once a CbfB-deficient mutant becomes available.

Preliminary data have suggested that the JmjC/ZF portion of CbfA is sufficient to rescue JH.D cell development in the absence of the CbfA CTD, which would imply that the two domains of the CbfA protein function independently (S. Gatz and T. Winckler, unpublished data). This finding would suggest that gene regulation mediated by the CbfA CTD depends on the direct binding to specific promoter regions in target genes; gene regulation mediated by the ectopically expressed CbfA CTD is compromised by mutations in the AT hook motif (18). It is tempting to speculate that the JmjC/ZF domain of CbfA acts as a chromatin modifier without directly binding to DNA. This possibility would assume that the CbfA zinc fingers have functional parallels with plant homeodomain (PHD) fingers, which are known to interpret the histone code by binding to nucleosome core histones in a modification-dependent manner (37, 38). CbfA displays significant similarity to the *Neurospora* DMM-1 protein in both the JmjC and ZF domains but not in the CTD. DMM-1 is required to inhibit the spreading of histone 3 lysine-9 methylation from transposable element loci into nearby genes (39). Notably, both the JmjC and ZF domains are required for the function of DMM-1, while the distinct CTD of DMM-1 is dispensable, even though the presence of multiple AT hooks in the DMM-1 CTD would have suggested a role for this region in directing the protein to its target sequences. In fact, DMM-1 requires interactions with the DMM-2 protein to interact with target DNA (39).

Rapid evolution of CTDs. In this study, we defined a new family of "C-module-binding factor-like proteins" that are characterized by the unique combination of a JmjC domain and CbfA-archetypical zinc fingers. CbfA-like proteins emerged before the dictyostelid split, and the dictyostelid clade is characterized by the emergence of a second CbfA-like protein, i.e., CbfA. Whereas CbfA-like proteins are characterized by the rapid evolution of the CTDs, the CbfA CTD is highly conserved and is detectable only in the dictyostelid clade. The question that remains is whether CbfA has diverged from a CbfB-like ancestor or vice versa. The differences in the ZF1 region in the CbfA and CbfB proteins suggest that the CbfB proteins represent the ancestral variant of the ZF1 region. This region was subsequently expanded to the dictyostelid-specific CbfA proteins by the addition of the CX_2C and CX_2H motifs (see Fig. S4 in the supplemental material). Assuming that the zinc finger(s) in the CbfA proteins may be involved in the binding to specific histone modifications as a part of an intrinsic chromatin regulatory function similar to the plant PHD fingers (37, 38), it is possible that adaptations in the ZF1 region have yielded functional specifications in the JmjC/ZF regions of the CbfA and CbfB proteins. This hypothesis can be evaluated by comparing the gene expression in CbfA- and CbfB-deficient mutants. We speculate that CbfA arose in the dictyostelid clade by the duplication of an ancestral CbfA-like protein that consisted of a CbfB-like JmjC/ZF region and an unknown CTD. The recent discovery of a CbfA-like protein within an *Acanthamoeba* protein (Glöckner, unpublished) supports the suggestion that CbfA derived from the dictyostelid lineage by gene duplication from a CbfB-like ancestor. The CbfA CTD may have been a part of the ancestral CbfA-like protein that has been maintained by evolutionary pressures to serve common gene regulatory functions. Whether the rapid divergence of the CTDs in CbfA-like proteins (including *D. discoideum* CbfB) occurred through a mutation in a preexisting CTD or by frequent domain swapping remains unknown. The extended CbfA CTDs in the earlier dictyostelids may mediate functions specialized to the individual properties of the respective genomes.

The data presented in this study suggest that evolution has combined two independent gene regulatory functions into a single protein; whereas one of the regulatory elements was notably well conserved (JmjC/ZF), the other (CTD) was highly variable. This variability enabled the adaptation to the high genome flexibility that occurred in dictyostelid evolution (4). We predict that, similar to the CbfA CTDs, the CbfB CTDs have autonomous gene regulatory functions, but these functions are specialized to individual genomes and do not exhibit much overlap between species. However, a complete loss of function of this domain is also possible. This hypothesis can now be further tested experimentally on the basis of the data presented in this paper.

ACKNOWLEDGMENTS

We are grateful to H. Urushihara of the University of Tsukuba and A. Kuspa of the Baylor College of Medicine for discussing data from comparative dictyostelid genome projects prior to publication. We thank P.

Schaap for depositing many of the dictyostelid strains required for this study at the *Dictyostelium* Stock Center, and we thank the Stock Center staff for providing these strains.

This work was supported by the German Research Foundation DFG (WI 1142/6-1 and WI 1142/7-1).

REFERENCES

1. Kessin RH. 2001. *Dictyostelium*—evolution, cell biology, and the development of multicellularity. Cambridge University Press, Cambridge, United Kingdom.
2. Bonner JT. 2009. The social amoebae: the biology of cellular slime molds. Princeton University Press, Princeton, NJ.
3. Eichinger L, Pachebat JA, Glöckner G, Rajandream M-A, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, Tunggal B, Kummerfeld S, Madera M, Konfortov BA, Rivero F, Bankier AT, Lehmann R, Hamlin N, Davies R, Gaudet P, Fey P, Pilcher K, Chen G, Saunders D, Sodergren E, Davis P, Kerhornou A, Nie X, Hall N, Anjard C, Hemphill L, Bason N, Farbrother P, Desany B, Just E, Morio T, Rost R, Churcher C, Cooper J, Haydock S, van Driessche N, Cronin A, Goodhead I, Muzny D, Mourier T, Pain A, Lu M, Harper D, Lindsay R, Hauser H, James K, Quiles M, Madan Babu M, Saito T, Buchrieser C, Wardroper A, Felder M, Thangavelu M, Johnson D, Knights A, Loulseged H, Mungall K, Oliver K, Price C, Quail MA, Urushihara H, Hernandez J, Rabbinowitsch E, Steffen D, Sanders M, Ma J, Kohara Y, Sharp S, Simmonds M, Spiegler S, Tivey A, Sugano S, White B, Walker D, Woodward J, Winckler T, Tanaka Y, Shaulsky G, Schleicher M, Weinstock G, Rosenthal A, Cox EC, Chisholm RL, Gibbs R, Loomis WF, Platzer M, Kay RR, Williams J, Dear PH, Noegel AA, Barrell B, Kuspa A. 2005. The genome of the social amoeba *Dictyostelium discoideum*. Nature 435:43–57.
4. Heidel AJ, Lawal HM, Felder M, Schilde C, Helps NR, Tunggal B, Rivero F, John U, Schleicher M, Eichinger L, Platzer M, Noegel AA, Schaap P, Glöckner G. 2011. Phylogeny-wide analysis of social amoeba genomes highlights ancient origins for complex intercellular communication. Genome Res. 21:1882–1891.
5. Sucgang R, Kuo A, Tian X, Salerno W, Parikh A, Feasley CL, Dalin E, Tu H, Huang E, Barry K, Lindquist E, Shapiro H, Bruce D, Schmutz J, Fey P, Gaudet P, Anjard C, Mohan MB, Basu S, Bushmanova Y, van der Wel H, Katoh M, Coutinho PM, Saito T, Elias M, Schaap P, Kay RR, Henrissat B, Eichinger L, Rivero-Crespo F, Putnam NH, West CM, Loomis WF, Chisholm R, Shaulsky G, Strassmann JE, Queller DC, Kuspa A, Grigoriev I. 2011. Comparative genomics of the social amoebae *Dictyostelium discoideum* and *Dictyostelium purpureum*. Genome Biol. 12: R20.
6. Glöckner G, Szafranski K, Winckler T, Dingermann T, Quail M, Cox E, Eichinger L, Noegel AA, Rosenthal A. 2001. The complex repeats of *Dictyostelium discoideum*. Genome Res. 11:585–594.
7. Hedges DJ, Deininger PL. 2007. Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. Mutat. Res. 616:46–59.
8. Huang CR, Burns KH, Boeke JD. 2012. Active transposition in genomes. Annu. Rev. Genet. 46:651–675.
9. Winckler T, Schiefner J, Spaller T, Siol O. 2011. *Dictyostelium* transfer RNA gene-targeting retrotransposons: studying mobile element-host interactions in a compact genome. Mob. Genet. Elements 1:145–150.
10. Beck P, Dingermann T, Winckler T. 2002. Transfer RNA gene-targeted retrotransposition of *Dictyostelium* TRE5-A into a chromosomal UMP synthase gene trap. J. Mol. Biol. 318:273–285.
11. Marschalek R, Brechner T, Amon-Böhm E, Dingermann T. 1989. Transfer RNA genes: landmarks for integration of mobile genetic elements in *Dictyostelium discoideum*. Science 244:1493–1496.
12. Schumann G, Zündorf I, Hofmann J, Marschalek R, Dingermann T. 1994. Internally located and oppositely oriented polymerase II promoters direct convergent transcription of a LINE-like retroelement, the *Dictyostelium* repetitive element, from *Dictyostelium discoideum*. Mol. Cell. Biol. 14:3074–3084.
13. Bilzer A, Dölz H, Reinhardt A, Schmith A, Siol O, Winckler T. 2011. The C-module-binding factor supports amplification of TRE5-A retrotransposons in the *Dictyostelium discoideum* genome. Eukaryot. Cell 10: 81–86.
14. Horn J, Dietz-Schmidt A, Zündorf I, Garin J, Dingermann T, Winckler T. 1999. A *Dictyostelium* protein binds to distinct oligo(dA) x oligo(dT) DNA sequences in the C-module of the retrotransposable element DRE. Eur. J. Biochem. 265:441–448.
15. Winckler T, Trautwein C, Tschepke C, Neuhauser C, Zündorf I, Beck P, Vogel G, Dingermann T. 2001. Gene function analysis by amber stop codon suppression: CMBF is a nuclear protein that supports growth and development of *Dictyostelium* amoebae. J. Mol. Biol. 305:703–714.
16. Dingermann T, Reindl N, Brechner T, Werner H, Nerke K. 1990. Nonsense suppression in *Dictyostelium discoideum*. Dev. Genet. 11:410–417.
17. Winckler T, Iranfar N, Beck P, Jennes I, Siol O, Baik U, Loomis WF, Dingermann T. 2004. CbfA, the C-module DNA-binding factor, plays an essential role in the initiation of *Dictyostelium discoideum* development. Eukaryot. Cell 3:1349–1358.
18. Lucas J, Bilzer A, Moll L, Zündorf I, Dingermann T, Eichinger L, Siol O, Winckler T. 2009. The carboxy-terminal domain of *Dictyostelium* C-module-binding factor is an independent gene regulatory entity. PLoS One 4:e5012. doi:10.1371/journal.pone.0005012.
19. Raper KB. 1984. The dictyostelids. Princeton University Press, Princeton, NJ.
20. Papadopoulos JS, Agarwala R. 2007. COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics 23:1073–1079.
21. Levi S, Polyakov M, Egelhoff TT. 2000. Green fluorescent protein and epitope tag fusion vectors for Dictyostelium discoideum. Plasmid 44:231–238.
22. Hentschel U, Zündorf I, Dingermann T, Winckler T. 2001. On the problem of establishing the subcellular localization of *Dictyostelium* retrotransposon TRE5-A proteins by biochemical analysis of nuclear extracts. Anal. Biochem. 296:83–91.
23. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara Catenazzi E M, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ng BL, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Pinkard DC, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Rodriguez AC, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Sohna Sohna JE, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, vandeVondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53–59.
24. Mortazavi A, Williams BA, McClue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5:621–628.
25. Team RDC. 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
26. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.
27. Anders S, Huber W. 2010. Differential expression analysis for sequence count data. Genome Biol. 11:R106. doi:10.1186/gb-2010-11-10-r106.

28. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, AmiGO Hub, Web Presence Working Group. 2009. AmiGO: online access to ontology and annotation data. Bioinformatics 25:288–289.
29. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876–4882.
30. Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504.
31. Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29:e45.
32. Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, Shi Y. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119:941–953.
33. Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P, Zhang Y. 2006. Histone demethylation by a family of JmjC domain-containing proteins. Nature 439:811–816.
34. Klose RJ, Kallin EM, Zhang Y. 2006. JmjC-domain-containing proteins and histone demethylation. Nat. Rev. Genet. 7:715–727.
35. Agger K, Christensen J, Cloos PA, Helin K. 2008. The emerging functions of histone demethylases. Curr. Opin. Genet. Dev. 18:159–168.
36. Schaap P, Winckler T, Nelson M, Alvarez-Curto E, Elgie B, Hagiwara H, Cavender J, Milano-Curto A, Rozen DE, Dingermann T, Mutzel R, Baldauf SL. 2006. Molecular phylogeny and evolution of morphology in the social amoebas. Science 314:661–663.
37. Taverna SD, Li H, Ruthenberg AJ, Allis CD, Patel DJ. 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat. Struct. Mol. Biol. 14:1025–1040.
38. Sanchez R, Zhou MM. 2011. The PHD finger: a versatile epigenome reader. Trends Biochem. Sci. 36:364–372.
39. Honda S, Lewis ZA, Huarte M, Cho LY, David LL, Shi Y, Selker EU. 2010. The DMM complex prevents spreading of DNA methylation from transposons to nearby genes in *Neurospora crassa*. Genes Dev. 24:443–454.

6.3 Manuskript 3

The *Dictyostelium discoideum* RNA-dependent RNA polymerase RrpC silences the centromeric retrotransposon DIRS-1 post-transcriptionally and is required for the spreading of RNA silencing signals

---

Stephan Wiegand, Doreen Meier, Carsten Seehafer, Marek Malicki, Patrick Hofmann, **Anika Schmith**, Thomas Winckler, Balint Földesi, Benjamin Boesler, Wolfgang Nellen, Johan Reimegard, Max Käller, Jimmie Hällman, Olof Emanualsson, Lotta Avesson, Frederik Söderbom und Christian Hammann

Publiziert in: *Nucleic Acids Research*, 2014

3330–3345 Nucleic Acids Research, 2014, Vol. 42, No. 5 Published online 24 December 2013
doi:10.1093/nar/gkt1337

The *Dictyostelium discoideum* RNA-dependent RNA polymerase RrpC silences the centromeric retrotransposon DIRS-1 post-transcriptionally and is required for the spreading of RNA silencing signals

Stephan Wiegand[1], Doreen Meier[2], Carsten Seehafer[1], Marek Malicki[1], Patrick Hofmann[1], Anika Schmith[3], Thomas Winckler[3], Balint Földesi[2], Benjamin Boesler[2], Wolfgang Nellen[2], Johan Reimegård[4], Max Käller[4], Jimmie Hällman[4], Olof Emanuelsson[4], Lotta Avesson[5], Fredrik Söderbom[6,7] and Christian Hammann[1,*]

[1]Ribogenetics@Biochemistry Lab, School of Engineering and Science, Molecular Life Sciences Research Center, Jacobs University Bremen, Campus Ring 1, DE-28759 Bremen, Germany, [2]Abteilung Genetik, Universität Kassel, Heinrich-Plett-Strasse 40, DE-34132 Kassel, Germany, [3]Friedrich-Schiller-Universität Jena, Institut für Pharmazie, Lehrstuhl für Pharmazeutische Biologie, Semmelweisstraße 10, DE-07743 Jena, Germany, [4]Division of Gene Technology, KTH Royal Institute of Technology, Science for Life Laboratory (SciLifeLab Stockholm), School of Biotechnology, SE-171 65 Solna, Sweden, [5]Garvan Institute of Medical Research, 384 Victoria St Darlinghurst, NSW 2010, Australia, [6]Department of Cell and Molecular Biology, Biomedical Center, Uppsala University, PO Box 596, S-75124 Uppsala, Sweden and [7]Science for Life Laboratory, SE-75124 Uppsala, Sweden

Received August 6, 2013; Revised November 29, 2013; Accepted November 30, 2013

ABSTRACT

Dictyostelium intermediate repeat sequence 1 (DIRS-1) is the founding member of a poorly characterized class of retrotransposable elements that contain inverse long terminal repeats and tyrosine recombinase instead of DDE-type integrase enzymes. In *Dictyostelium discoideum*, DIRS-1 forms clusters that adopt the function of centromeres, rendering tight retrotransposition control critical to maintaining chromosome integrity. We report that in deletion strains of the RNA-dependent RNA polymerase RrpC, full-length and shorter DIRS-1 messenger RNAs are strongly enriched. Shorter versions of a hitherto unknown long non-coding RNA in DIRS-1 antisense orientation are also enriched in *rrpC*$^-$ strains. Concurrent with the accumulation of long transcripts, the vast majority of small (21 mer) DIRS-1 RNAs vanish in *rrpC*$^-$ strains. RNASeq reveals an asymmetric distribution of the DIRS-1 small RNAs, both along DIRS-1 and with respect to sense and antisense orientation. We show that RrpC is required for post-transcriptional DIRS-1 silencing and also for spreading of RNA silencing signals. Finally, DIRS-1 mis-regulation in the absence of RrpC leads to retrotransposon mobilization. In summary, our data reveal RrpC as a key player in the silencing of centromeric retrotransposon DIRS-1. RrpC acts at the post-transcriptional level and is involved in spreading of RNA silencing signals, both in the 5′ and 3′ directions.

INTRODUCTION

Cellular RNA-dependent RNA Ppolymerases (RdRPs) are present in many eukaryotic organisms, including mammals (1), are involved in mechanisms of RNA-mediated gene regulation including RNA interference (RNAi) (2,3). RdRPs synthesize an RNA strand from a complementary RNA template, leading to either long or small antisense transcripts. Based on *in vivo* and *in vitro* studies, including deep sequencing analyses, several primer-dependent and -independent modes of action have been inferred for RdRPs from different organisms. These frequently lead to an amplification of a primary silencing signal such as double-strand-derived small interfering RNA (siRNA). An amplifying component

*To whom correspondence should be addressed. Tel: +49 4212 003 247; Fax: +49 4212 003 249; Email: c.hammann@jacobs-university.de

was postulated by Fire and Mello when they first described RNAi in *Caenorhabditis elegans* (4), leading to the identification of RdRP EGO-1 as the responsible enzyme (5). When RNA was injected against a reporter gene, secondary siRNAs were observed that predominantly localize 5′ of the original trigger, and this 5′ spreading was shown to be RdRP-dependent (6). This phenomenon, known as 'transitivity', is diagnostic for RdRP activity *in vivo* (6 8). We recently observed transitivity in the amoeba *Dictyostelium discoideum*, and have shown that it depends on the presence of the RdRP RrpC and the dicer-related nuclease DrnB, as inferred from an artificial reporter system (9). RrpC is one of three RdRPs in *D. discoideum*, all of which show a similar domain structure (Supplementary Figure S1) (10).

Dictyostelium intermediate repeat sequence 1 (DIRS-1) is the founding member of a poorly characterized class of long terminal repeat (LTR) retrotransposons. DIRS elements differ from other LTR retrotransposons by having inverted instead of direct terminal repeats, lacking an aspartic protease domain and using a tyrosine recombinase instead of a DDE-type integrase protein for integration. Thus, DIRS elements use a retrotransposition mechanism fundamentally different from that of other LTR retrotransposons [reviewed in (11)]. This may favor the recombinatorial integration into preexisting copies of the same element as seen in the *D. discoideum* genome. The genome of *D. discoideum* features ~40 intact copies and ~200 300 fragments of DIRS-1, which thus represents the most frequently occurring LTR retrotransposon in the amoeba (12). DIRS-1 sequences accumulate at one end of each chromosome (12), and these clusters have been suggested to represent the centromeres of the chromosomes in *D. discoideum* (13,14). Although observed clustering of DIRS-1 elements may result from preferential integration of mobilized DIRS-1 copies into preexisting DIRS-1 clusters (15), it is not known whether the apparent clustering of these elements is the result of deleterious integration into coding regions of the haploid *D. discoideum* genome that removes affected cells from the population (16). Nonetheless, uncontrolled amplification of DIRS-1 elements and integration into centromeric regions may seriously compromise centromere stability, and thus genome integrity.

DIRS-1 is transcribed as a 4.5-kb-long messenger RNA (mRNA) with three overlapping open reading frames (ORFs) (Figure 1A) (17). The left LTR possesses promoter activity to drive the transcription of the DIRS-1 mRNA (18). These DIRS-1 mRNA transcripts were found to accumulate during *D. discoideum* development and contain parts of the two LTRs (18). Additionally, heat shock or other stress conditions can trigger the expression of a 900-nt-long antisense transcript (Figure 1A), which, however, is not expressed under the axenic growth conditions applied in this study (19,20).

For retrotransposition to occur, the mRNA is thought to be reverse transcribed by the enzyme activity of ORF III (17). The ends of the resulting linear complementary DNA (cDNA) feature the LTR fragments, which exhibit sequence complementarity to a DIRS-1 region termed the internal complementary region (ICR). According to a model put forward by Lodish *et al.* the ICR serves to bring the 5′- and the 3′-end of the reverse transcript together (11) to generate a single-stranded circular DNA replication intermediate that serves as template for the generation of a double-stranded circular DNA to be inserted in the genome (17).

We have shown earlier that the LTR of DIRS-1 is methylated by the sole DNA methyltransferase, DnmA, identified in *D. discoideum* (21). Surprisingly, no transcriptional activation was observed in a *dnmA* gene deletion strain, indicating that DIRS-1 is not under transcriptional control, at least not by DnmA. However, in that study, we observed that regulation of DIRS-1 expression may be controlled at the post-transcriptional level involving small RNAs (21). Here, we show that RrpC controls DIRS-1 post-transcriptionally, and its activity is required to prevent DIRS-1 retrotransposition. We further report RrpC-dependent spreading of an RNA silencing signal in the 3′ direction in *D. discoideum*.

MATERIALS AND METHODS

Oligonucleotides

DNA oligonucleotides (Sigma) used in this study are listed in Supplementary Table S1.

Genes and strains

All experiments were carried out with AX2 and derivatives. All *D. discoideum* strains were grown axenically in HL5 medium. The *rrp* genes are listed in the online resource dictybase.org with their accession numbers DDB_G0289659 (*rrpA*), DDB_G0291249 (*rrpB*) and DDB_G0280963 (*rrpC*), and the generation of their single and multiple deletion strains was described recently (9). The *trx-1* gene has the accession number DDB_G0294447.

RNA isolation

Total RNA was isolated from 5×10^7 cells of axenically grown *D. discoideum* strains using TRIzol® (Invitrogen). Cells were pelleted for 5 min at 500 rcf/4°C and washed in 20 ml pre-cooled Sorenson phosphate buffer (2 mM Na_2HPO_4, 15 mM KH_2PO_4, pH 6.0, with H_3PO_4). On re-centrifugation of cells under the same conditions, they were lysed in 1 ml of TRIzol® reagent for 5 min. On addition of 200 µl of chloroform, the material was vortexed and incubated for 3 min at room temperature. Phases were separated by centrifugation at 16 000 rcf/4°C for 20 min. RNA was precipitated by addition of equal amounts of 100% isopropanol, incubation for 10 min at room temperature and centrifugation at 16 000 rcf/4°C for 15 min. On washing twice with 70% ethanol, the pellet was dried and re-suspended in 150 µl dH_2O. RNA concentration was determined using a NanoDrop spectrophotometer (Peqlab).

Northern blotting

For total RNA blots, 10 µg was separated by gel electrophoresis in a 1.2% GTC-agarose gel in 1 × Tris

borate + ethylenediaminetetraacetic acid (pH 8.2). On capillary transfer to nylon membranes (Roti®-Nylon plus), ultraviolet cross-linking was carried out (0.5 J/cm^2). Prehybridization and hybridization were carried out in Church buffer [500 mM sodium phosphate (pH 7.2), 1 mM ethylenediaminetetraacetic acid, 7% w/v sodium dodecyl sulphate (SDS), 1% w/v bovine serum albumin] using radioactively labeled *in vitro* transcripts or polymerase chain reaction (PCR) products. Hybridization conditions are summarized in Supplementary Tables S2 and S3. All blots were washed twice for 15 min in 2 × saline sodium citrate (SSC), 0.1% w/v SDS, twice for 10 min in 1 × SSC, 0.1% w/v SDS and twice for 5 min in 0.5 × SSC, 0.1% w/v SDS. As control, blots were hybridized on stripping to probes against actin6 mRNA.

For small RNA blotting, 20μg total RNA was separated by gel electrophoresis in a 11% polyacrylamide gel in 20 mM MOPS (pH 7.0). On electroblotting onto a nylon membrane (Amersham Hybond^{TM}-NX) for 10 min at 20 V in dH_2O (semi-dry), RNA was chemically cross-linked as described recently (22). Small RNAs were probed with 5′ radioactively labeled DNA oligonucleotides as summarized in Supplementary Table S4. All other procedures were carried out as described for ultraviolet cross-linking, except for the use of the small nucleolar RNA (snoRNA) DdR-6 as loading control.

Southern blotting

Isolation of genomic DNA from 1 2×10^8 *Dictyostelium* cells for Southern blotting (23) was performed as described (24). For DNA digestion with restriction enzymes, aliquots containing 20 μg of DNA were incubated with *Sac* I, *Kpn* I or *Eco* RI (20 U/10 μg of DNA). After precipitation, samples were separated on 1.2% agarose gels for 14 h at a field strength of 1.5 V/cm, and fragments were visualized by ethidium bromide staining (0.5 μg/ml) for 30 min. DNA was denatured by soaking the gel in 10 volumes of denaturation buffer (1.5 M NaCl, 0.5 M NaOH) for 45 min, followed by 45 min soaking in neutralization buffer (3 M NaCl, 0.5 M Tris-HCl, pH 7). DNA fragments were transferred to a nylon membrane as described (25). Cross-linking and hybridization were carried out as described for northern blotting, using the right element sequence, labeled with [α-^{32}P]dATP by random priming. On overnight incubation, the membrane was washed with 3 × SSC, 0.01% SDS and then with 0.1 x SSC, 0.1% SDS, each at 60°C and for 1 h.

Generation of expression constructs

pTX-DIRS-1rLTRfw-GFP was generated by initially digesting the original vector pTX-GFP (26) with *Sal* I and ligating the oligonucleotides pTX-MCS-fw and pTX-MCS-rev for generating a multiple cloning site (MCS). Depending on the orientation of the MCS, the resulting vectors were termed pTX-MCSfw-A15-GFP and pTX-MCSrev-A15-GFP. To insert the DIRS-1 rLTR-sequence, plasmid pGEM®-T Easy DIRS-1 rLTR (21) was digested with *Sac* II and *Sal* I. Ligation into the MCS of pTX-MCSfw-A15-GFP and pTX-MCSrev-A15-GFP yielded the plasmids pTX-DIRS-1rLTRfw-A15-GFP and pTX-DIRS-1rLTRrev-A15-GFP. The former was digested with *Sac* II and *Xho* I to isolate the expression cassette, which was ligated into the vector pLPBLP (27), that had been digested with the same enzymes beforehand. From this, the Actin15-promoter was removed by a restriction digest with *Hin*d III and *Xho* I. The 5′ overhangs were filled by Klenow fragment before ligation, and the expression cassettes without Actin15-promoter were ligated back in the original vector using a *Sac* II and *Xho* I, which yielded the plasmid pTX-DIRS-1rLTRfw-GFP. From pTX-DIRS-1rLTRrev-A15-GFP, the expression cassette was removed with *Sal* I and *Bam*H I and ligated into the vector pJET1-H2Bv1, from which the H2Bv1 fragment had been removed with the same enzymes. The actin15-promoter was removed by a digest with *Hin*d III and *Sac* II, and the vector was re-ligated on removal of 5′ overhangs using Klenow fragment. The expression cassette was back-cloned into the original vector using *Sal* I and *Bam*H I, yielding plasmid pTX-DIRS-1rLTRrev-GFP.

The DIRS-1 ORFs were amplified by PCR using primers listed in Supplementary Table S1. PCR products were ligated into the cloning pJET1.2/blunt, and the resulting plasmids were sequenced. The ORF DNAs were excised by means of *Bam*H I and *Spe* I digests and ligated into pDM323 (28) that was previously linearized with *Bgl* II and *Spe* I. The resulting plasmids were transformed in AX2 and rrpC knockout strains, and transformation populations were used for further experiments. For the GFP-alone control, plasmid pDM317 (28) was used that features an ATG in front of the GFP coding sequence, allowing to monitor GFP expression by western blotting.

Fluorescence *in situ* hybridization and microscopy

Fluorescence *in situ* hybridization (FISH), using digoxigenin-labeled dNTPs, and 4,6-diamidino-2-phenylindole (DAPI) staining were performed as described recently (29). For the generation of probes for FISH, oligonucleotides DIRS1 forward and DIRS1 reverse were used, resulting in a 503-nt-long single-stranded DNA of DIRS-1 sense or antisense orientation, depending on their relative concentration in asymmetric PCRs. The fragment begins at the left end (LE) of ORF III (Figure 1) and is extended by 200 nt. Pictures were generated using an Alexa Fluor 488-coupled secondary antibody.

Quantitative real-time PCR

Expression analysis by quantitative real-time PCR (qRT-PCR) was carried out as described recently (30,31). In brief, total RNA was isolated using the Qiagen RNeasy Mini kit according to the provided protocol from 2×10^7 cells that were collected from exponentially growing cells and frozen at −80°C until further use. cDNA was synthesized by reverse transcription of 500 ng of total RNA using an oligodeoxythymidine primer and the Qiagen Omniscript RT kit. The Q1 and Q2 fragments of DIRS-1 were amplified from oligo(dT)-primed cDNA

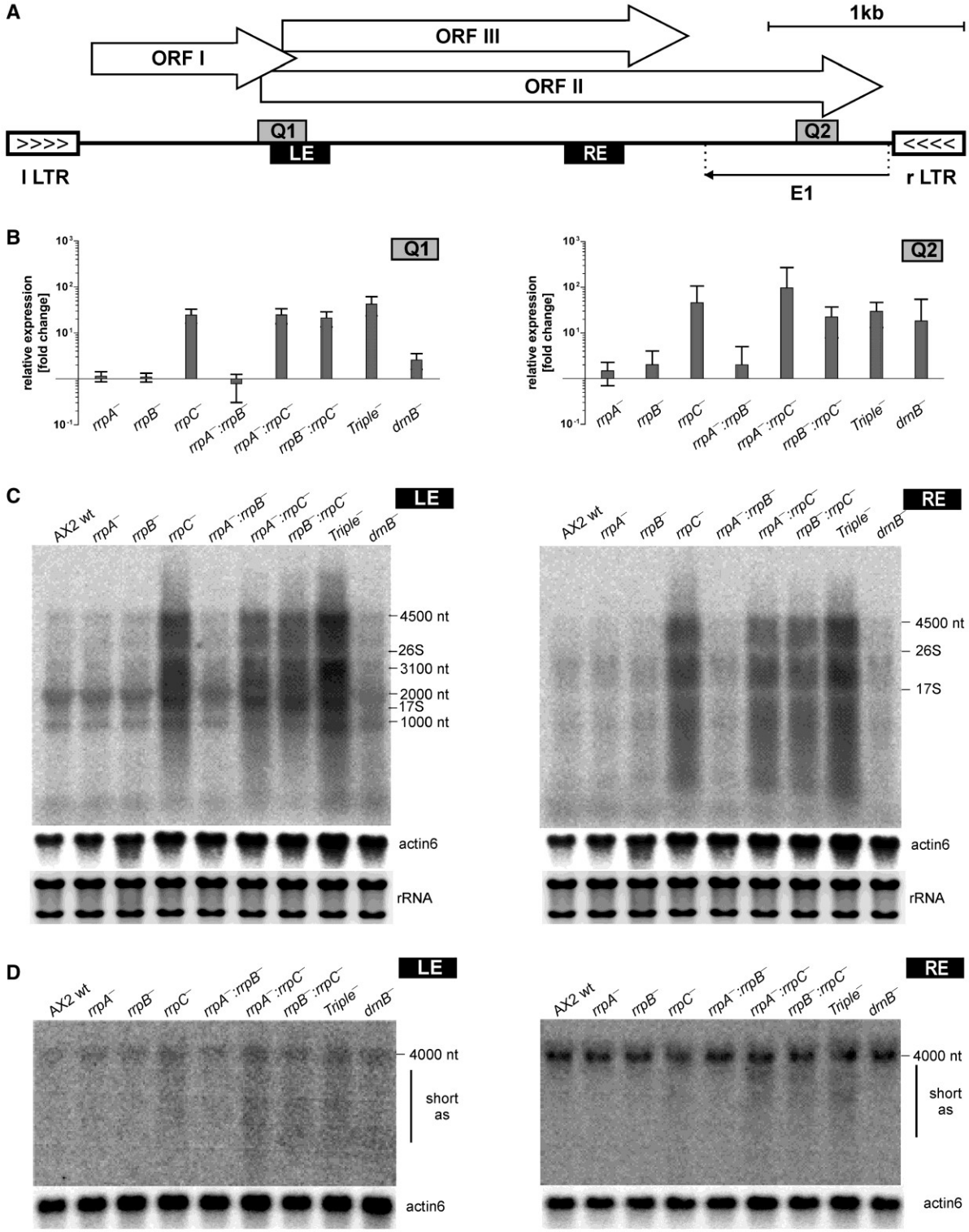


Figure 1. DIRS-1 expression. (**A**) Schematic representation of the DIRS-1 retrotransposon with the left and right inverted LTRs (l LTR and r LTR, respectively). The three ORFs, encoding the GAG protein (ORF I), the tyrosine recombinase (ORF II) and reverse transcriptase/RNase H domain and a methyltransferase (ORF III) are displayed together with the 900-nt E1 antisense transcript. Positions for expression analysis are Q1 and Q2 for qRT-PCR and LE and RE for northern blots. (**B**) Expression of DIRS-1 sequences in the indicated gene deletion strains, relative to the AX2 wild-type, as monitored by qRT-PCR using primer pairs at positions Q1 and Q2, and normalized to GAPDH expression. (**C**) Northern blot analysis of RNA from the indicated strains, using DIRS-1 sense strand-specific probes at positions LE (left) and RE (right). Ethidium bromide-stained rRNA served as loading control, while in northern blots, an actin6 probe acted to monitor transfer efficiency. (**D**) Northern blots using DIRS-1 antisense strand-specific probes at positions LE (left) and RE (right). Other details are as in (C). Size ranges of signals for shorter antisense RNAs are indicated to the right of each blot. Northern blot conditions are listed in Supplementary Tables S2 and S3. See also Figures 2 and 3 and Supplementary Figure S2.

with primer pairs Q-DIRS-01/-02 and Q-DIRS-03/-04, respectively. Expression of DIRS-1 was standardized against expression of the GAPDH gene, which was amplified using primers Q-gpdA-01 and Q-gpdA-02. Real-time amplification was carried out on a Stratagene Mx3000P instrument in 25 µl reaction mixes containing 1 × Taq buffer, 0.2 mM deoxynucleoside triphosphates, 0.4 µM primers, 1 × ROX reference dye (Jena Bioscience, Germany), 1 × EvaGreen solution (Jena Bioscience, Germany), varying amount of cDNA and 1.25 U Taq polymerase (Jena Bioscience, Germany). After an initial denaturing step at 95°C for 10 min, the PCR was performed for 40 cycles at 95°C for 30 s, 58°C for 30 s and 72°C for 30 s. Regulation was calculated by the method of Pfaffl (32).

Deep sequencing and computational analysis

RNA was prepared and converted to cDNA, from which sequencing libraries were prepared and sequenced using the Illumina platform as described recently (33). Endogenous siRNA populations in the AX2 wild-type and *rrpC* knockout strains, each in two biological replicates, were sequenced after ligation of Illumina TruSeq small RNA adapters that allow cluster amplification on the flow cell surface. Adapter sequences were trimmed from the sequences using Cutadapt (34) with at least 3-nt overlap allowing for 10% mismatches. The trimmed reads were mapped against a reference using Bowtie2 (35). The reference was either the genome of *D. discoideum* strain AX4 (downloaded from http://dictybase.org/) or our own repeat sequence library containing simple repeats from Repbase (36) and consensus sequences of interspersed repeats including DIRS-1. The default setting was used when mapping to the genome reference, while a more sensitive setting was used when mapping to the repeat reference (-M 500 -N 1 -L 20 -R 3 -D 20 -i S,1,0.50). Read count per annotated region (http://dictybase.org/download/gff3/) was extracted with HTSeq (http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html). The read distribution was visualized with the integrative genomics viewer (37).

RESULTS

Accumulation of DIRS-1 sense transcripts in the absence of the *rrpC* gene

In a previous study, we observed the accumulation of transcripts of the retrotransposon DIRS-1 in a gene disruption strain of the RdRP RrpC (21). To study the impact of RdRPs on DIRS-1 expression in *D. discoideum* in detail, we made use of a recently introduced series of strains (9), with the genes of the RdRPs in all possible combinations deleted by a one-step cloning strategy (38) that is based on the Cre-lox system (27). Unlike in the originally used gene disruption strains (21), both functional domains of each *rrp* gene (Supplementary Figure S1) were removed completely (9), resulting in three single, three double and one triple *rrp* gene deletion strain. In addition, DIRS-1 expression was monitored in a gene deletion strain of the dicer-related nuclease DrnB (*drnB*$^-$) (39). Analyses by qRT-PCR using the two primer sets Q1 and Q2 (Figure 1A) showed a 50 100-fold increase of DIRS-1 RNA in all *rrpC* gene deletion strains compared with the isogenic wild-type strain (Figure 1B). In the *drnB*$^-$ strain, a 5 10-fold increase in RNA levels was observed, depending on which primer set was used. To assess how these RNA molecules correspond to full-length DIRS-1 RNA, we next carried out northern blot analyses with riboprobes against the left-end (LE) and right-end (RE) positions in the DIRS-1 sense transcript that localize to the 5′-end or the 3′ part of ORF III, respectively (Figure 1A). As shown in Figure 1C, these analyses revealed a considerable increase of a 4.5-kb fragment corresponding to the size of the DIRS-1 full-length mRNA (18) for each *rrpC* gene deletion strain, while the transcript levels in the *drnB*$^-$ strain appear similar to that observed in the wild-type strain (Figure 1C). In the *rrpC*$^-$ strains, the DIRS-1 transcripts display strong size heterogeneity as indicated by the smear in the respective lanes of the northern blots (summarized in Table 1).

A long DIRS-1 antisense transcript

Considering that antisense transcripts are prevalent, particularly in mobile genetic elements (40,41), we also used riboprobes that bind to the same LE and RE positions (Figure 1A) on an antisense transcript. We observed an RNA transcript of 4 kb by either riboprobe (Figure 1D). The existence of this hitherto unknown antisense RNA was also confirmed independently by strand-specific qRT-PCR (Supplementary Figure S2). We termed this long non-coding RNA (lncRNA) *long antisense DIRS-1* (lasDIRS-1) to discriminate it from the 900-nt E1 antisense transcript that is produced on heat shock (19,20), but not under the axenic growth conditions used in this study. Unlike the situation seen for the DIRS-1 mRNA, the lasDIRS-1 lncRNA displays a discrete size, and its RNA levels appear to be identical in either gene deletion strain investigated here. Only in the *rrpC*$^-$ strains was a light smear below the full-length lasDIRS-1 RNA observed reproducibly, indicating size heterogeneity (Figure 1D and summarized in Table 1).

Promoter activity of the DIRS-1 right LTR

Because the sense transcript is derived from the promoter activity of the left LTR (18), it seemed plausible that the inversely oriented right LTR sequence might contain the promoter to drive lasDIRS-1 transcription. To test this hypothesis, we generated two constructs, which contained the right LTR in forward or reverse orientation in front of a GFP expression cassette, but lacked any further promoter sequence (Figure 2A). Fluorescence microscopy of strains transformed with these two constructs revealed GFP expression only for the construct with the forward-oriented right LTR (Figure 2B). The promoter activity of the right LTR of DIRS-1 was also independently confirmed by western blotting (Figure 2C). These data strongly suggest that the lasDIRS-1 transcript derives from the right LTR of the retrotransposon.

Table 1. Characteristics of rrp gene deletion strains

rrp gene deletion	Change in DIRS-1 transcripts levels			Spreading of small RNA signal	
	mRNA[a]	lasDIRS1[b]	Small RNAs[c]	In 5′ direction[d]	In 3′ direction[e]
rrpA^{-}	Unchanged	Unchanged	Unchanged	1	1
rrpB^{-}	Unchanged	Unchanged	Unchanged	1	1
rrpC^{-}	Strongly increased	Shorter versions increased	Strongly decreased	Not observed	Not observed
rrpA^{-}:*rrpB*$^{-}$	Unchanged	Unchanged	Unchanged	1	n.d.[f]
rrpA^{-}:*rrpC*$^{-}$	Strongly increased	Shorter versions increased	Strongly decreased	Not observed	n.d.[f]
rrpB^{-}:*rrpC*$^{-}$	Strongly increased	Shorter versions increased	Strongly decreased	Not observed	n.d.[f]
rrpA^{-}:*rrpB*$^{-}$:*rrpC*$^{-}$ (*Triple*$^{-}$)	Strongly increased	Shorter versions increased	Decreased	Not observed	n.d.[f]

[a]Relative to AX2 wild-type, as determined by qRT-PCR and northern blotting (Figure 1), [b]Relative to AX2 wild-type, as determined by qRT-PCR and northern blotting (Figure 1), [c]Relative to AX2 wild-type, as determined by northern blotting (Figure 4), [d]Shown in an earlier study using a β-galactosidase reporter assay (9), [e]As determined by northern blotting (Figure 6).
f n.d., not determined.

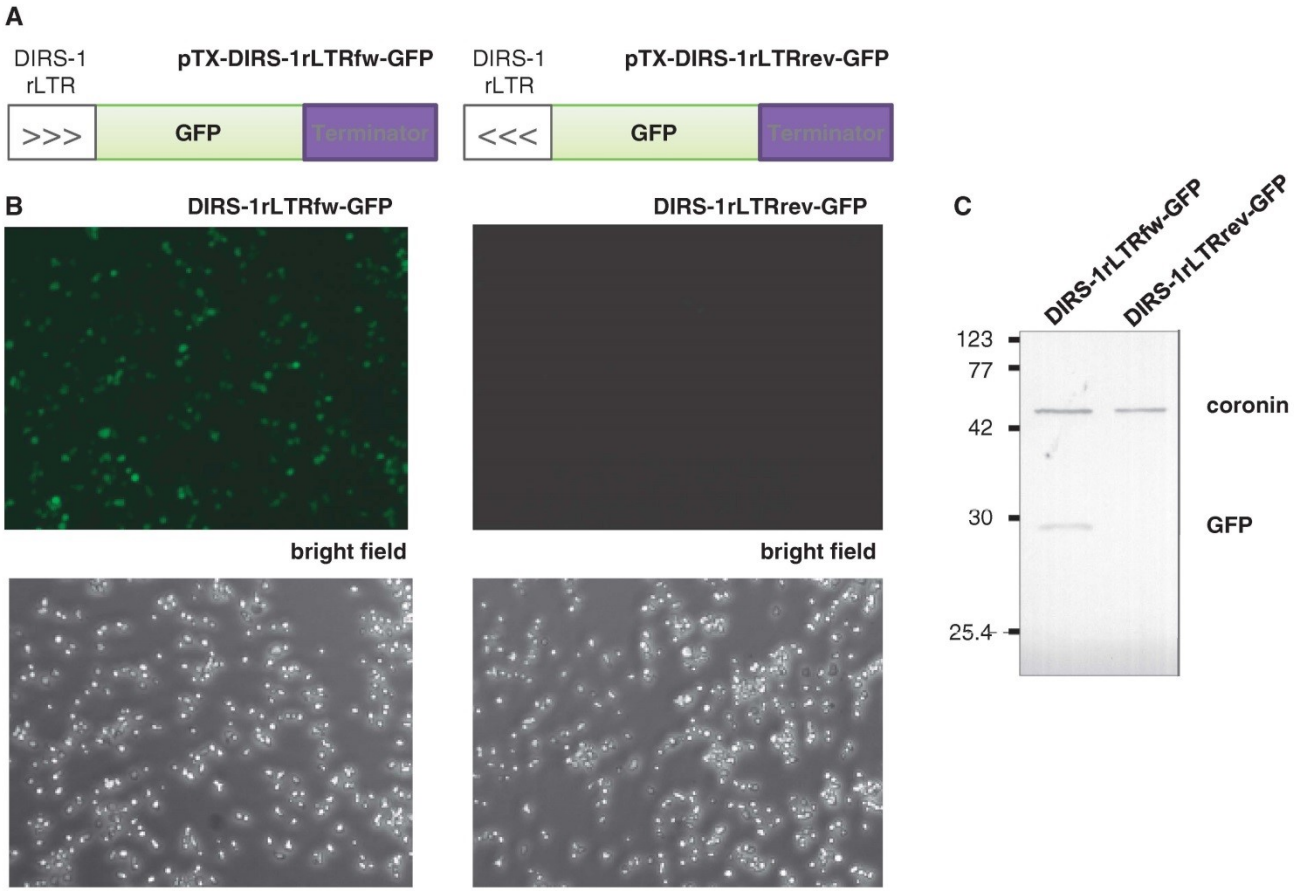


Figure 2. Analysis of DIRS-1 right LTR promoter activity. (**A**) Constructs with forward and reverse right LTR, a GFP ORF and terminator sequence. (**B**) Fluorescence (top) and bright field (bottom) microscopy of AX2 cells transformed with constructs shown in (A). (**C**) Western blot of these transformed cells using antibodies against GFP and Coronin (loading control). The molecular weights of a size marker (in kDa) are indicated to the left.

DIRS-1 transcripts cluster in the nucleus of *rrpC*$^{-}$ cells

To determine the localization of DIRS-1 transcripts in amoeba cells, we used a recently established protocol for RNA-FISH in *D. discoideum* (29). First, we generated digoxigenin-labeled strand-specific 500-nt-long probes starting at the LE position of DIRS-1 (Figure 1A), which we used in hybridization experiments of fixed and permeabilized AX2 wild-type and *rrpC*$^{-}$ cells. Fluorescence microscopy revealed for either probe diffuse signals that predominantly localized to the cytoplasm in AX2 wild-type cells, as judged from the DAPI staining of the nucleus (Figure 3A). In the *rrpC*$^{-}$ strain,

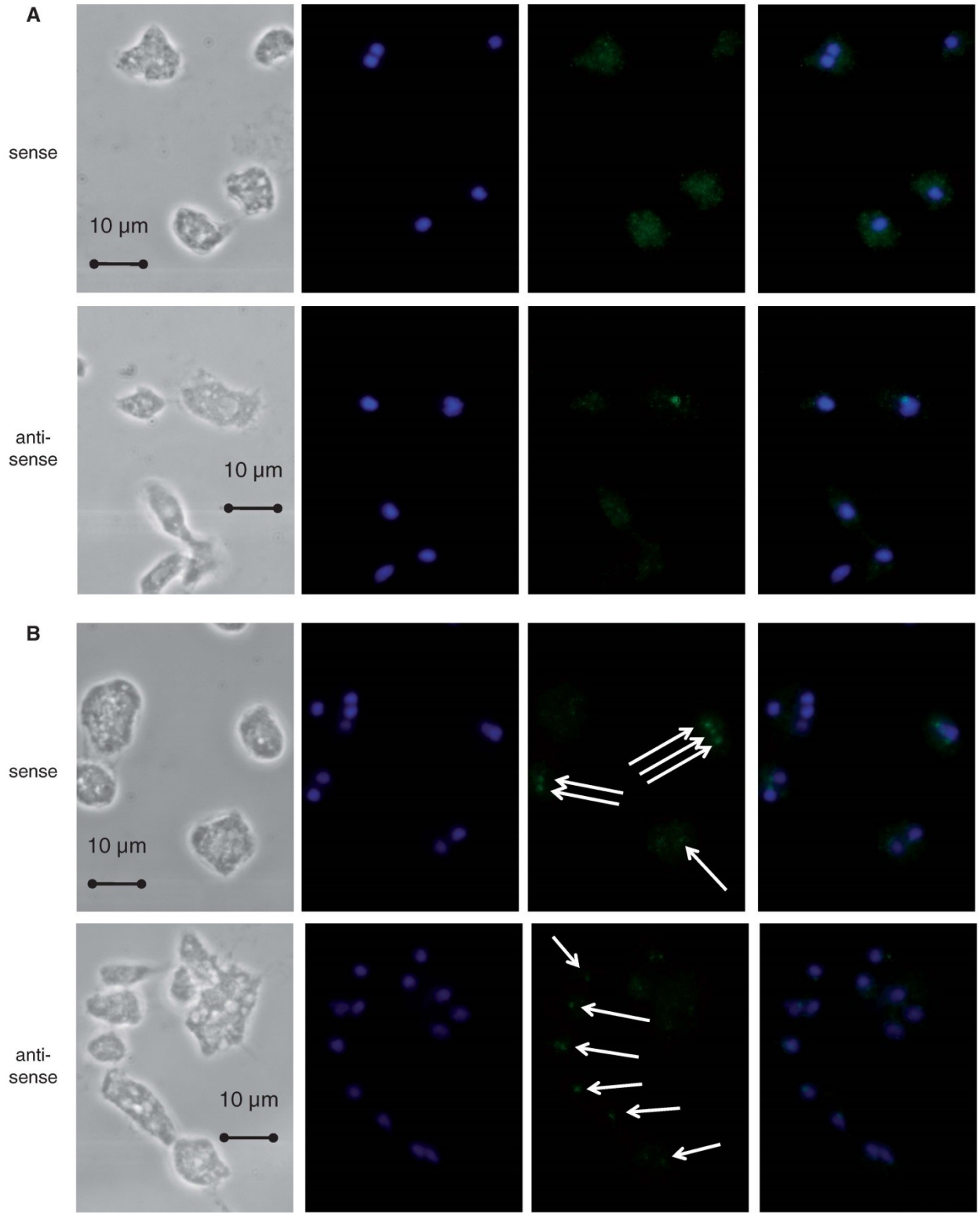


Figure 3. Localization of DIRS-1 transcripts in *D. discoideum* cells by FISH. Sense and antisense DIRS-1 transcripts in AX2 (**A**) and *rrpC*$^{-}$ cells (**B**). Shown are microscopic images of fixed and permeabilized cells. First column: phase contrast; second column: DAPI staining; third column: Alexa Fluor 488-coupled secondary antibody staining of DIRS-1 transcripts; and fourth column: merge of DAPI and Alexa Fluor 488-coupled secondary antibody staining. Arrows mark accumulated transcripts. All pictures were taken at a 100-fold magnification, and a scale bar is indicated for each row in the phase contrast panel. Pictures were taken with exposure times of 260–350 ms for DAPI and 700–1000 ms for Alexa.

additional strong signals were observed (Figure 3B), which appeared to cluster in one or more spots in the nucleus. These clusters of DIRS-1 transcripts were observed by using either the sense or the antisense probe in the *rrpC*$^{-}$ strain (Figure 3B).

As summarized in Table 1, our data indicate that full-length as well as shorter sense DIRS-1 RNAs accumulates in *rrpC*$^{-}$ strains. A long antisense RNA, lasDIRS-1, is also expressed and likely driven by the promoter activity of the right LTR. Shorter fragments of lasDIRS-1 were found to accumulate in the *rrpC*$^{-}$ strains. Compared with the AX2 wild-type strain, additional signals for DIRS-1 sense and antisense transcripts appeared to localize in nuclear spots in the *rrpC*$^{-}$ strain.

DIRS-1-specific small RNAs

Because gene silencing by RdRPs is commonly associated with the appearance of small RNAs, we next analyzed the presence of DIRS-1-specific small RNAs by northern blotting. Using a radioactively labeled oligonucleotide corresponding to the LE position of the retrotransposon, we observed a significant reduction of DIRS-1-specific small RNAs in all strains lacking the *rrpC* gene as compared with the wild-type (Figure 4A). The reduction is also apparent, albeit to a lesser extent, in a northern blot using a probe against the RE position of DIRS-1 (Figure 4B). The individual deletions of the other *rrp* genes, *rrpA* or *rrpB* appeared to have no significant effect on the accumulation of these DIRS-1 small RNAs, and the same is observed for DrnB (Figure 4).

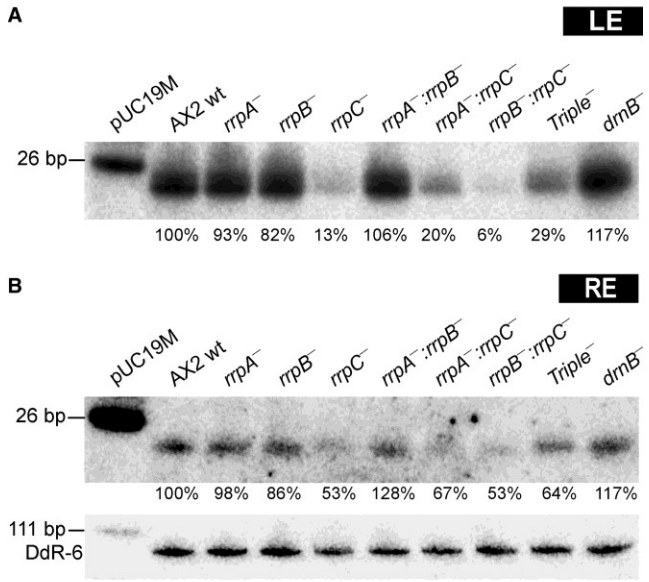


Figure 4. DIRS-1-specific small RNAs expression. Northern blots monitoring the small RNAs of DIRS-1 in the indicated strains, using oligoprobes LE (**A**) or RE (**B**), which are localized to DIRS-1 as indicated in Figure 1A. Quantification of small RNA levels (indicated below the blots) is relative to that of the AX2 wild-type and normalized to the DdR-6 snoRNA sequence (42). The 26-bp band of the size marker pUC19 DNA/*Msp*I is indicated. Northern blot conditions are listed in Supplementary Table S4.

However, with both probes, the decrease appeared less pronounced in the triple knockout strain (see also Table 1), an observation that we cannot explain at present. Yet, these northern blot data indicate that the RdRP RrpC is required for the generation of DIRS-1 small RNAs, at least at the positions LE and RE (Figure 1A). In our previous study, we had observed small RNAs distributed over the entire sequence of the retrotransposon (21). To analyze whether all these DIRS-1 small RNAs depend on RrpC, we next carried out deep sequencing studies on RNA from *D. discoideum* AX2 wild-type and *rrpC*$^{-}$ cells.

Asymmetric distribution of small RNAs over the DIRS-1 sequence

The small RNA profiles of AX2 wild-type and *rrpC*$^{-}$ strains were determined by deep sequencing on an Illumina platform using two biological replicates per strain. This resulted in ~28 and ~16 million sequence reads of 21 nt for AX2 and *rrpC*$^{-}$, respectively. These reads were mapped against the *D. discoideum* genome and a library of repeat sequences (manuscript in preparation).

In the wild-type AX2 strain, a substantial fraction of small RNA sequencing reads (19.5%) corresponded to DIRS-1 sequences, confirming our earlier studies that used 5′-ligation-dependent cloning and SOLiD sequencing of small RNAs from the AX4 strain (43). However, in the *rrpC*$^{-}$ strain, the number of DIRS-1 sequencing reads is strongly (−65%) reduced from 5.5 to 1.9 M, indicating that RrpC might be involved in their generation. We next mapped these small RNA reads on the DIRS-1 sequence (Figure 5A). The comparison of the small RNA distribution patterns in the two biological replicates of the wild-type AX2 strain (Figure 5B) indicates that the data are highly reproducible, and the same holds for the two biological replicates of the *rrpC*$^{-}$ strain (Figure 5C).

The distribution of small RNAs over the DIRS-1 sequence appears uneven in the AX2 strain. The majority of reads coincide with ORF II, while the 5′-end of GAG appears to be devoid of any small RNAs (Figure 5B). This analysis also showed the reduction of the small DIRS-1 RNAs in the *rrpC*$^{-}$ strain, but additionally indicated that not all positions within the retrotransposon are equally affected (Figure 5C). For example, although reduced reads are observed for the majority of DIRS-1 positions in the *rrpC*$^{-}$ strain, this is not the case for the relatively small number of reads that map to the LTRs of the retrotransposon (box I, Figure 5B and C). These LTR-specific small RNAs appear to be unchanged in number and position and thus likely are independent of RrpC. A small number of reads corresponding to the 5′-end of GAG was surprisingly observed only in the *rrpC*$^{-}$ strain, but not in the AX2 wild-type (box II, Figure 5B and C). These small RNAs thus seem to disappear when RrpC is present. Reads mapping around the 3′-end of ORF I and the 5′-end of ORF II appear to be particularly reduced in the *rrpC*$^{-}$ strain (box III, Figure 4B and C), while the distribution of the majority of small RNAs in the *rrpC*$^{-}$ strain has an overall similar

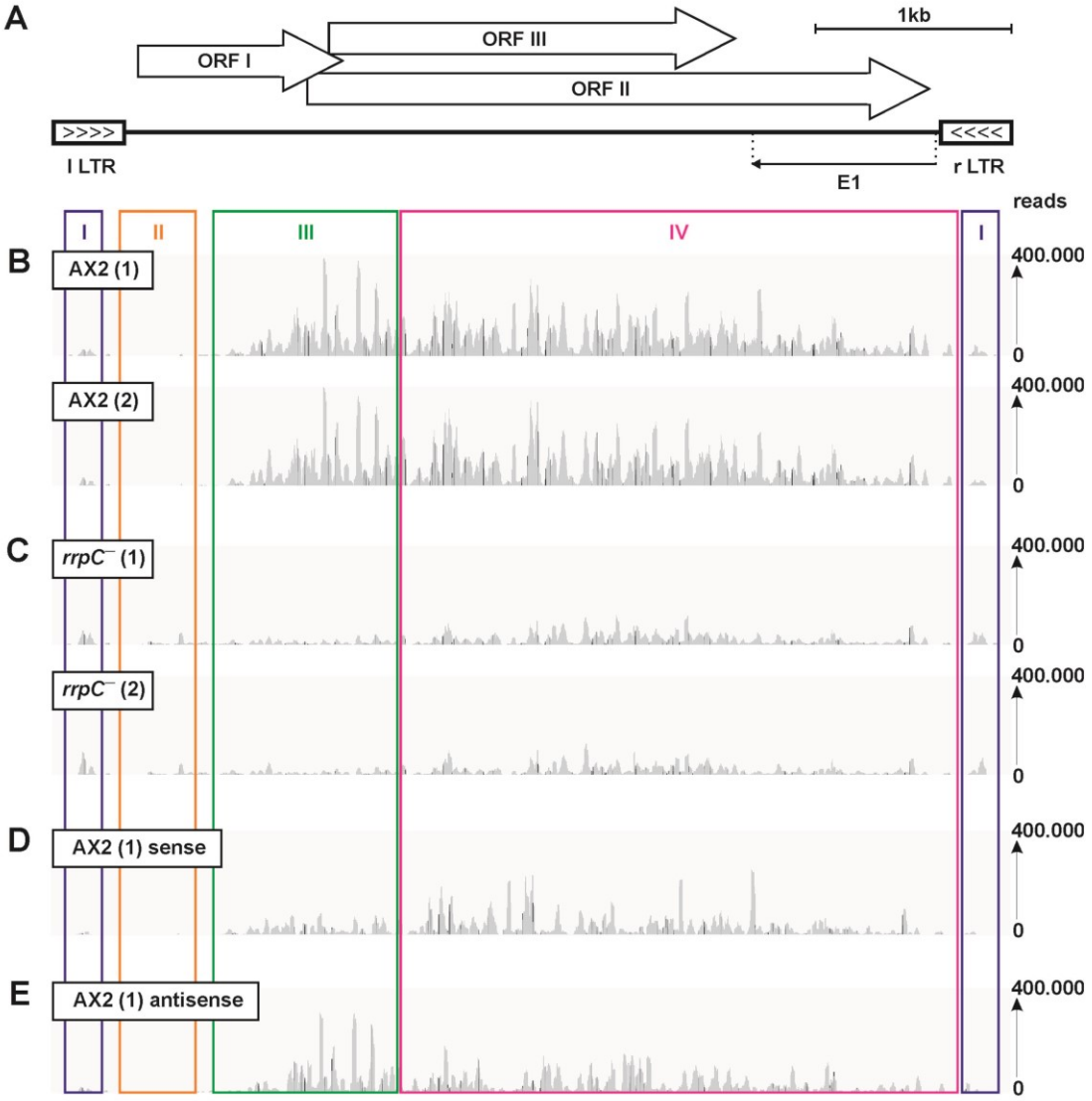


Figure 5. Distribution of DIRS-1-specific small RNAs. (**A**) Schematic representation of the DIRS-1 sequence (for details see Figure 1A). Small RNA reads from the wild-type AX2 strain (**B**) and the *rrpC*⁻ strain (**C**) are shown for their biological duplicates (1 or 2, respectively). The DIRS-1 small RNAs of sample (1) are shown separately for sense (**D**) and antisense (**E**) orientations. All reads are mapped onto the DIRS-1 sequence in (A). The signal height corresponds to the number of small RNAs, and reads are displayed in the same scale, ranging from 0 to 400.000 reads. Regions with particularly apparent deviations between the AX2 and *rrpC*⁻ strains are boxed (I, II, III, IV).

pattern to that observed in the AX2 wild-type strain, albeit at drastically reduced numbers (box IV, Figure 5B and C). These observations suggest that both of these classes of DIRS-1 small RNAs depend on RrpC.

Asymmetric distribution of DIRS-1 small RNAs in sense and antisense orientation

In view of the presence of the lasDIRS-1 transcript (Figures 1D and 3), we next analyzed whether the DIRS-1-specific small RNAs mapped equally in sense and antisense orientation to the retrotransposon. The results showed uneven distribution of the small RNAs, with the sense small RNAs mapping predominantly to ORF II and the antisense RNAs appearing to peak in the region box III (Figure 5D and E). The asymmetry of sense and antisense read distribution indicates that the majority of small DIRS-1 RNAs might not be generated as classical double-stranded siRNAs.

Post-transcriptional silencing of DIRS-1 mRNAs

Because the small DIRS-1 RNAs appear to be unevenly distributed along the three ORFs of the retrotransposon, we wondered what effect they might have on these three ORFs. To investigate this question, we generated over-expression constructs, in which GFP was C-terminally fused to the full mRNA of either ORF (Supplementary Figure S3). The resulting extrachromosomal vector constructs were transformed in the AX2 wild-type and the

rrpC¯ strain. Expression of the fusion proteins was monitored by western blotting using a GFP antibody. The ORF I-GFP fusion protein was readily expressed in both strains, while expression of the other two ORFs was not detectable in the AX2 wild-type, but notably observed in the *rrpC*¯ strain (Figure 6A and B). This indicates that the RrpC-dependent small RNAs that are present in the AX2 wild-type strain are necessary and

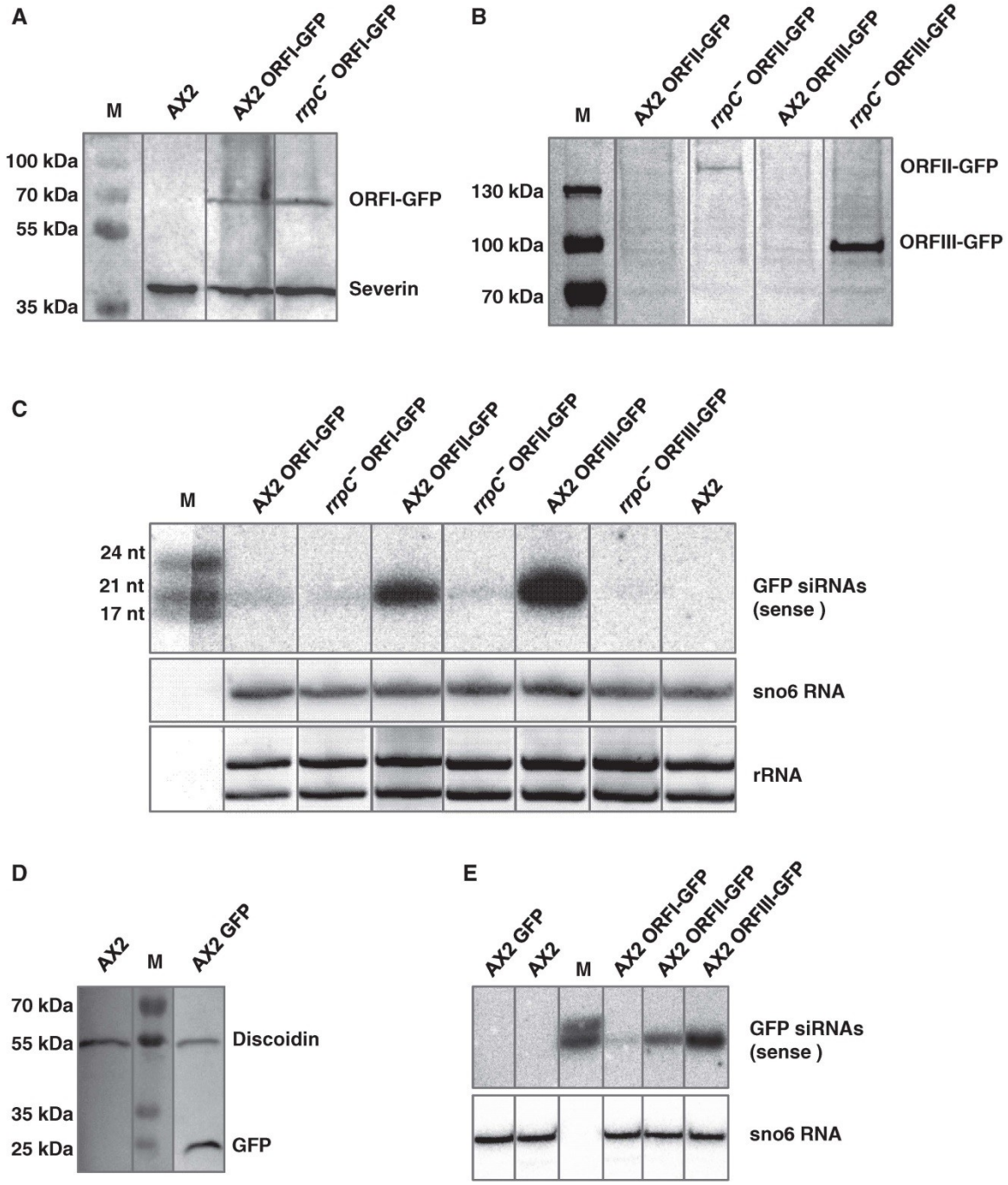


Figure 6. Overexpression of DIRS-1 ORFs. C-terminal GFP fusions of DIRS-1 ORF I (**A**) and ORF II or ORF III (**B**) were expressed in the AX2 wild-type and in the *rrpC*¯ strain. The expression of the fusion proteins was monitored by western blotting using a GFP antibody (see also Supplementary Figure S3). (**C**) Small GFP RNAs from the indicated strains were analyzed by northern blot using a strand-specific radioactively labeled probe. The sizes of a radioactively labeled RNA size marker are shown to the left. Ethidium bromide-stained rRNA served as loading control, and northern blots using a sno6 riboprobe acted to monitor transfer efficiency. (**D**) Expression of GFP in AX2 wild-type cells. (**E**) Small GFP RNAs from the indicated strains. In (**A**, **B** and **D**), severin or discoidin were used as loading controls, and the sizes of a protein marker are shown. M in (**C** and **E**) denotes small RNA size markers. Northern blot conditions are listed in Supplementary Table S4. Note that boxed lanes are spliced together from the same blot deleting irrelevant samples.

sufficient to silence post-transcriptionally ORF II and ORF III, but not ORF I, of the retrotransposon, even though siRNAs in the 3′ half of ORF I are mostly antisense to the coding transcript and could readily abolish it.

Spreading of the RNA silencing signal in the 3′ direction on the target mRNA

We have recently shown that RrpC is required for the appearance of small RNAs 5′ of the original silencing signal with respect to the target RNA (9). This phenomenon is generally referred to as transitive silencing (6 8), and it is thought that the target RNA serves as the template, from which an RdRP synthesizes a complementary RNA strand. Because of the directionality of all nucleic acid polymerases, such a secondary silencing signal is expected to localize 5′ of the original trigger. However, Pak and Fire reported that a subset of secondary RNAs could also localize 3′ of the original trigger in *C. elegans* (44), an observation that otherwise is mainly seen in the plant kingdom (7,8,45,46). To test whether 3′ spreading might also be observed in the case of the DIRS-1 ORF-GFP fusion constructs, we isolated small RNAs from these strains and used northern blot analysis to identify small RNAs derived from GFP (Figure 6C). Substantial amounts of small GFP-specific sense RNAs were detected in the ORF II-GFP and ORF III-GFP expressing cells in the AX2 wild-type background, but not in the *rrpC*$^{-}$ strain. To show that GFP itself does not cause the generation of siRNAs, we transformed the AX2 wild-type with a GFP expression construct alone, which resulted in GFP expressing cells (Figure 6D) and did not cause generation of small GFP RNAs as monitored by northern blotting (Figure 6E). The inclusion of RNA from the C-terminal GFP fusions of the DIRS-1 ORFs serves here as control for the functionality of the blot, as it otherwise would appear empty. Because the GFP was fused C-terminally to the DIRS-1 ORFs, this result indicates that spreading of an RNA silencing signal in the 3′ direction exists in *D. discoideum* and depends on the presence of RrpC (summarized in Table 1). We note that a weak signal for small GFP RNAs is also present in strains where the fusion protein is not silenced, but not in the untransformed AX2 wild-type strain (Figure 6).

Mobilization of DIRS-1 in the genome of the *rrpC*$^{-}$ strain

Having shown with the ORF GFP fusions that DIRS-1 mRNAs are translated in the *rrpC*$^{-}$ strain, we next set out to analyze what effects this has on the genome of the *rrpC*$^{-}$ strain. However, the genome-wide analysis of DIRS-1 elements is hampered by the large number of DIRS-1 elements and fragments (12) and the fact that DIRS-1 retrotransposition takes place preferentially in other DIRS-1 elements or fragments (15).

To investigate whether new DIRS-1 integrations can be observed in the genome of the *rrpC*$^{-}$ strain, we subjected *rrpC*$^{-}$ and AX2 wild-type strains to extensive growth for 48 days in shaking cultures kept in the exponential growth phase by regular dilution. Under such conditions, we had earlier observed increased genomic copy numbers of Skipper, when that retrotransposon was dysregulated in a gene deletion strain of the putative DNA methyltransferase DnmA (21). From the AX2 and *rrpC*$^{-}$ long-term cultures, genomic DNA was isolated and digested with three different restriction enzymes of which *Eco* RI and *Kpn* I cleave inside the DIRS-1 DNA (Figure 7A), whereas *Sac* I cleaves outside. The restriction patterns of genomic DNA from these two strains appeared to be highly similar (Figure 7B). In the subsequently performed Southern blot (Figure 7C), the overall picture is also similar, as might be expected given the many genomic occurrences of DIRS-1 elements. For the *Eco* RI and the *Sac* I digests, the most prominent signals from the AX2 strain appear to be more intense in the *rrpC*$^{-}$ strain genomic DNA, which may indicate an increase in DIRS-1 copy number. More importantly, we observe an additional signal at ~5 kb in the *rrpC*$^{-}$ strain in the *Kpn* I digest, clearly indicating a DIRS-1 retrotransposition event (Figure 7C). To quantify these differences in DIRS-1 copy number, Southern blotting was repeated and signals for DIRS-1 and the thioredoxin gene as loading control were quantified for two digests of genomic DNA from AX2 wild-type and the *rrpC*$^{-}$ strain (Supplementary Table S5). This analysis confirmed an increase in DIRS-1 copy numbers by >50% in the *rrpC*$^{-}$ strain compared with the AX2 wild-type, as shown by loading-corrected signal increases of 52% and 62% for the two restriction enzymes used (Supplementary Table S5). Together, the data indicate that additional genomic copies of the retrotransposon accumulate in the genome of the *rrpC*$^{-}$ strain. This is further supported by a previous analysis of DIRS-1 copy numbers in the originally generated gene disruption strain of the *rrpC* gene (21), where different DIRS-1 patterns were observed in clonal long-term cultures (M. Kuhlmann and W. Nellen, unpublished).

DISCUSSION

In the present study, we show that the RdRP RrpC acts in the regulation of the centromeric retrotransposon DIRS-1. We report that sense transcripts accumulate in strains lacking the *rrpC* gene (Figure 1; summarized in Figure 8) and concurrent with this accumulation, DIRS-1-specific small RNAs are dramatically reduced (Figures 4 and 5; summarized in Table 1). As a consequence of their absence, it is likely that the DIRS-1 ORFs II and III are no longer silenced, leading to the accumulation of DIRS-1 proteins as monitored by GFP fusions (Figure 6) and DIRS-1 retrotransposition (Figure 7) in the *rrpC*$^{-}$ cells. From these data, we propose that DIRS-1 is constitutively transcribed in wild-type cells, and these transcripts are targeted by RrpC, which thus exerts it control predominantly on the post-transcriptional level.

For the full-length DIRS-1 elements, transcription is driven by the inherent promoter activity of the two LTRs, as reported earlier for the left LTR (18) and in this study for the right LTR (Figure 2), resulting in the newly discovered lncRNA antisense transcript lasDIRS-1 (Figure 1). The observation that the full-length lasDIRS-1

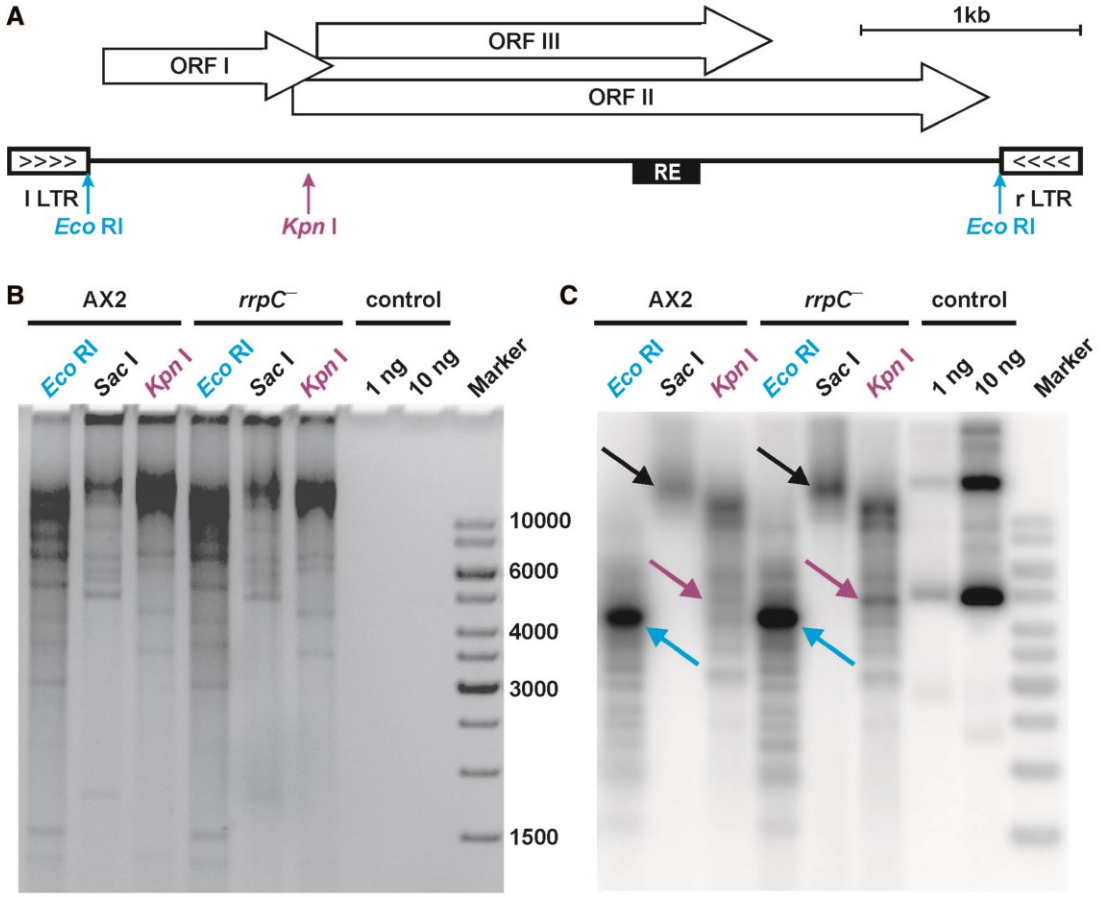


Figure 7. Southern blot analysis for genomic mobilization of DIRS-1 in the *rrpC*⁻ strain. (**A**) Schematic representation of DIRS-1 with recognition sites of *Eco* RI and *Kpn* I that cleave within the retrotransposon sequence, while *Sac* I (not shown) does not. (**B**) Agarose gel electrophoresis of 20 µg genomic DNA from AX2 wild-type and *rrpC*⁻ strains, digested with the indicated restriction endonucleases. As controls for hybridization, 1 or 10 ng undigested plasmid DNA featuring the DIRS-1 sequence was applied. The sizes of the 1-kb marker fragments (ThermoScientific) are indicated. (**C**) Southern blot using a ^{32}P-labeled DNA probe corresponding to the RE position. Note that the positions with an additional or stronger signals observed for the *rrpC*⁻ genomic DNA are indicated by arrows using the same colors for *Eco* RI and *Kpn* I as in (A). Changes in patterns of DNA digested with *Sac* I, which cleaves outside DIRS-1, are denoted by black arrows.

transcript is also present in the triple *rrp* gene deletion strain excludes the possibility that it might be an RdRP product. The sense and antisense transcripts that result from the LTR promoter activities appear to have different fates in the investigated cell lines. The levels of the lncRNA lasDIRS-1 are stable, whereas those of the sense RNA vary significantly (Figure 1; summarized in Figure 8). In particular, we observe a dramatic accumulation of full-length and shorter DIRS-1 sense transcripts, next to shorter versions of the lasDIRS-1 lncRNA in all deletion strains of the *rrpC* gene (Figure 1). As shown by RNA FISH, additional DIRS-1 signals can be observed in the nucleus of *rrpC*⁻ cells (Figure 3), and it appears plausible that the accumulating shorter sense and antisense transcripts that we observe by northern blotting are contained in these spots. Bidirectional transcription has also been observed for retrotransposons in other organisms, including LINE-1 in humans (47); LINE elements, SINE elements and LTR retrotransposons in mice (48,49); Tc1 in *C. elegans* (50); and transposons in *Drosophila* (51,52).

Concurrent with the accumulation of long transcripts, we observe by northern blot analyses and by deep sequencing (Figures 4 and 5), a drastic reduction of small DIRS-1 RNAs in strains lacking the *rrpC* gene. In the deep sequencing analysis, we find the majority of small RNAs to be asymmetrically distributed between the sense and antisense orientations of the retroelement in AX2 cells. This result indicates that these small RNAs are not reaction products of one of the two Dicer-like proteins in *D. discoideum* (21,39,43), because Dicers are expected to give rise to double-stranded siRNAs (53–55), which should cover DIRS-1 symmetrically in sense and antisense orientation.

Because the majority of small DIRS-1 RNAs disappear in the absence of RrpC, the simplest explanation is that these are directly synthesized by this RdRP. The synthesis of small RNAs by RdRPs has also been described in other species, including *C. elegans* (44,56), *Neurospora crassa* (57) and *Arabidopsis thaliana* (58). Our conclusion is also in line with our recently reported discovery that RrpC is required for the 5′ spreading of an RNA silencing

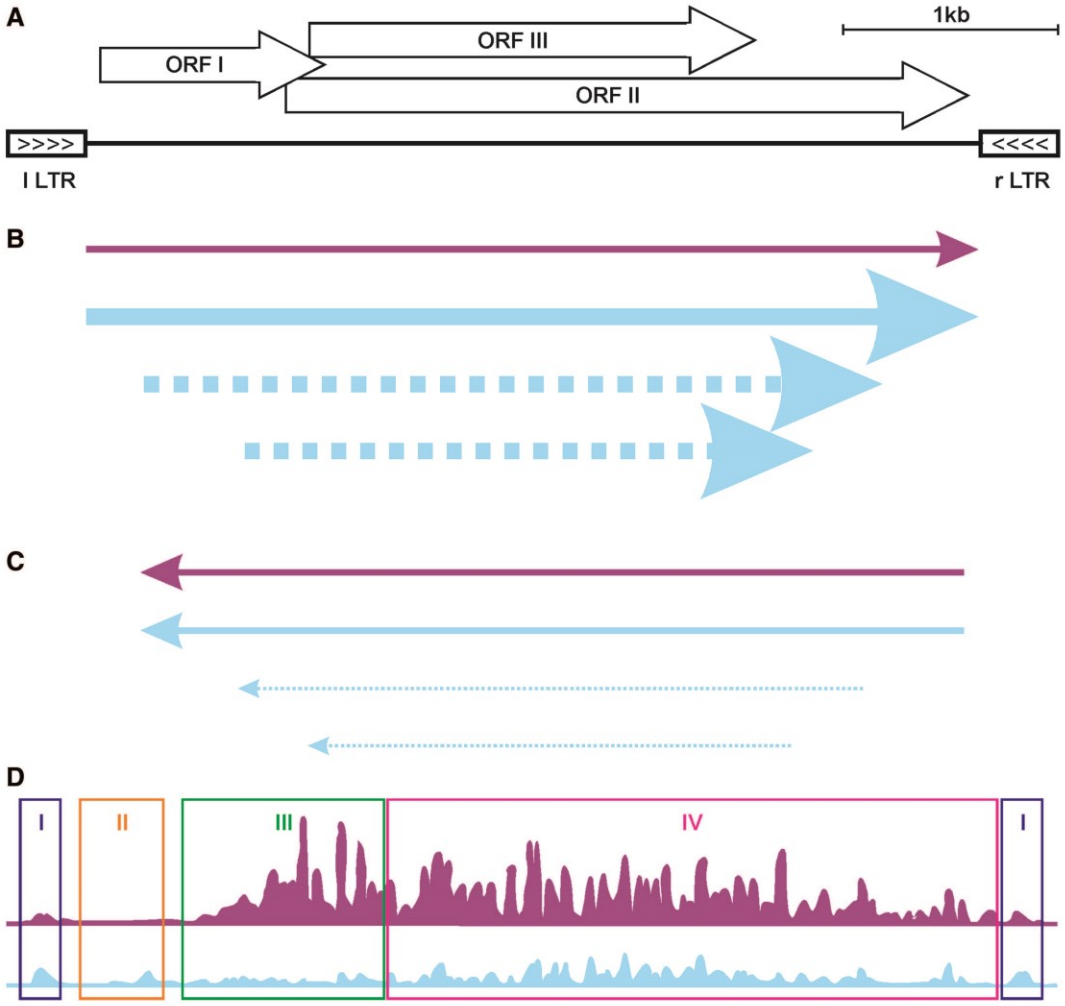


Figure 8. Summary of DIRS-1 RNAs in AX2 and *rrpC*¯ strains. (**A**) Schematic representation of the DIRS-1 sequence (for details see Figure 1A). Relative levels of DIRS-1 sense transcripts (**B**), antisense transcripts (**C**) and small RNAs (**D**) are shown for the AX2 wild-type (purple) and for the *rrpC*¯ (light blue) strains. Arrows in (B) and (C) represent the transcripts mapped onto the DIRS-1 element in (A). The thickness of the arrows indicates relative changes in RNA quantity. Note that RNA levels are comparable only within panels and that the positions of the shorter fragments (dotted) are not exactly determined. In (D), the distribution of small RNAs is also mapped onto the element in (A) and the height of the peaks indicates abundance, using the same relative scaling (for details see Figure 5).

signal (9). Changes in the distribution of small DIRS-1 RNAs in the *rrpC*¯ strain compared with AX2 appear to cluster in four regions (boxes I-IV in Figure 5; summarized in Figure 8), which correspond to the LTR with unchanged levels (I), to the 5′-end of GAG with small RNAs absent in AX2 (II), to a region with barely any small RNAs left (III) and to a region with reduced numbers but a similar distribution pattern (IV). For the latter, RrpC might serve as the amplifying component of a primary silencing signal, like a double-stranded RNA. Such an entity has been proposed earlier by Fire and Mello, when they discovered RNAi in *C. elegans* (4). The responsible enzyme was characterized subsequently as the RdRP EGO-1 (5). The small RNAs in region III appear to strongly accumulate in antisense orientation (Figure 5E). We propose that RrpC synthesizes these small RNAs, using the DIRS-1 sense transcript as a template. Why these RNA appear to be non-functional in the post-transcriptional silencing of the ORF I-GFP fusion protein (Figure 6A) is currently unclear, but one possible reason is that their concentration in that region might be too low (Figure 5).

The small RNAs of the LTR region, on the other hand, appear to be unaffected by the deletion of *rrpC*, and in view of their symmetric sense and antisense distribution, they might represent Dicer products. In this respect, it is worth noting that the sense transcript has been shown to contain little of the left LTR, but nearly the entire right LTR (18). Of these expressed right and left LTR sequences, 58 nt and 33 nt, respectively, are complementary to a DIRS-1 region termed the ICR. During DIRS-1 retrotransposition, the ICR is proposed to bring the 5′-end and the 3′-end of a reverse transcript together (11), to allow for the formation of a single-stranded circular DNA replication intermediate (17). It seems plausible to suggest that such base pairing also takes

place on the RNA level, and the resulting RNA double strands would be long enough to allow for Dicer activity, giving rise to the RrpC-independent DIRS-1 LTR siRNAs (box I in Figures 5 and 8).

Although the small DIRS-1 RNAs that are present only in the absence of RrpC (box II, Figures 5 and 8) are low in number, they appear reproducibly. Similarly, elevated small RNA levels in strains lacking the functional *rrpC* gene have also been shown for endogenous microRNAs (39,43), the retrotransposon Skipper (43), and recently also in transgene silencing (9). Given that RdRPs are enzymes that synthesize RNA, small RNA accumulation in their absence is counter-intuitive. However, some RdRPs in other organisms (57,58) have been proposed to act in a small RNA primer-dependent mode, in which the RNA primer is extended, resulting in the formation of long double-stranded RNAs, which act as *bona fide* Dicer substrates. If primer extension does not take place, the primers might accumulate. In view of small RNAs from various genetic backgrounds accumulating in *rrpC*$^{-}$ strains, a primer-dependent activity of this RdRP on these sequences appears to be a plausible interpretation. Albeit indirectly, this interpretation is also supported by the observation that classical (3′ 5′) transitive silencing in *D. discoideum* depends on both, the dicer-related nuclease DrnB and RrpC (9). Also for several other RdRPs, both primer-dependent and -independent modes of action have been inferred, like QDE-1 in *N. crassa* (57) or the *A. thaliana* enzyme RDR6 (58,59), which act in the primer-dependent mode also in the biosynthesis of tasiRNAs (58,60).

To our surprise, we observed significant levels of small RNAs 3′ to the endogenous DIRS-1 small RNA trigger during the analysis of the C-terminal GFP fusion constructs of the DIRS-1 ORFs (Figure 6). Substantial amounts of GFP-specific small RNAs were only detectable for the ORF II and ORF III fusion constructs, but not for ORF I-GFP. Early work on the activity of RdRPs has shown that unprimed activity is preferentially observed at the 3′-end of mRNAs (61,62). Because the small GFP RNAs are not observed in the GAG GFP construct, it appears unlikely that the GFP mRNA sequence itself is recognized as aberrant and is causative for the observed spreading in the 3′ direction. Because the presence of GFP alone does not lead to the generation of small GFP RNAs (Figures 2 and 6D and E), this possibility can be excluded. From the use of probes monitoring small GFP RNAs in the sense orientation, we conclude that these are not direct products of RrpC on the mRNA of the respective ORF GFP. Rather, they are likely processing products of an RrpC-dependant double-strand that might be targeted by one of the *D. discoideum* Argonaute or Dicer proteins.

In a previous study, we had investigated transcriptional silencing of DIRS-1 by DNA methylation. Although the DIRS-1 right LTR was found to be methylated by the sole DNA methyltransferase in *D. discoideum*, DnmA (21), a deletion of its gene did not result in a transcriptional upregulation of DIRS-1. From this, we concluded that DIRS-1 is not under transcriptional control by DnmA, unlike the retrotransposon Skipper (21). However, we cannot exclude the possibility that other components of the heterochromatization machineries in the amoeba might add a transcriptional control of DIRS-1 in addition to the post-transcriptional control by RrpC that we report here.

Control at the post-transcriptional level appears uneconomic, as DIRS-1 transcripts are first generated in order to then be degraded. However, this control might be a consequence of the integration behavior of DIRS-1. Because DIRS-1 retrotransposition takes place preferentially into other DIRS-1 elements or fragments thereof (15), highly diverse DIRS-1 sequences are expected to be produced. In these diverse DIRS-1 sequences, motifs other than the LTRs might serve as promoters, which then might render the mobile genetic elements refractory to transcriptional control, at least by using LTR promoter methylation.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Dianna K. Bautista and Timothy J. Wilson for most helpful comments on the manuscript. The authors acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure (NGI) and Uppmax for providing assistance in massively parallel sequencing and computational infrastructure. They also thank Markus Maniak for providing antibodies against GFP, Severin and Coronin.

FUNDING

The Deutsche Forschungsgemeinschaft [Heisenberg stipend HA3459/5-2, HA3459/7-1 to C.H., Wi1142/6-1 to T.W.]; by the European Molecular Biology Organisation [ASTF497-2011 to C.S.] and by the Swedish Research Council [to O.E. and F.S.]; Supported by a stipend from the Land Hessen (to D.M. and B.B.). Funding for open access charge: Deutsche Forschungsgemeinschaft.

Conflict of interest statement. None declared.

REFERENCES

1. Maida,Y., Yasukawa,M., Furuuchi,M., Lassmann,T., Possemato,R., Okamoto,N., Kasim,V., Hayashizaki,Y., Hahn,W.C. and Masutomi,K. (2009) An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA. *Nature*, **461**, 230–235.
2. Maida,Y. and Masutomi,K. (2011) RNA-dependent RNA polymerases in RNA silencing. *Biol. Chem.*, **392**, 299–304.
3. Wassenegger,M. and Krczal,G. (2006) Nomenclature and functions of RNA-directed RNA polymerases. *Trends Plant Sci.*, **11**, 142–151.
4. Fire,A., Xu,S., Montgomery,M.K., Kostas,S.A., Driver,S.E. and Mello,C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**, 806–811.

5. Smardon,A., Spoerke,J.M., Stacey,S.C., Klein,M.E., Mackin,N. and Maine,E.M. (2000) EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Curr. Biol.*, **10**, 169–178.
6. Sijen,T., Fleenor,J., Simmer,F., Thijssen,K.L., Parrish,S., Timmons,L., Plasterk,R.H. and Fire,A. (2001) On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell*, **107**, 465–476.
7. Vaistij,F.E., Jones,L. and Baulcombe,D.C. (2002) Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. *Plant Cell*, **14**, 857–867.
8. Voinnet,O., Vain,P., Angell,S. and Baulcombe,D.C. (1998) Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. *Cell*, **95**, 177–187.
9. Wiegand,S. and Hammann,C. (2013) The 5′ spreading of small RNAs in *Dictyostelium discoideum* depends on the RNA-dependent RNA polymerase RrpC and on the dicer-related nuclease DrnB. *PLoS One*, **8**, e64804.
10. Martens,H., Novotny,J., Oberstrass,J., Steck,T.L., Postlethwait,P. and Nellen,W. (2002) RNAi in Dictyostelium: the role of RNA-directed RNA polymerases and double-stranded RNase. *Mol. Biol. Cell*, **13**, 445–453.
11. Poulter,R.T. and Goodwin,T.J. (2005) DIRS-1 and the other tyrosine recombinase retrotransposons. *Cytogenet. Genome Res.*, **110**, 575–588.
12. Eichinger,L., Pachebat,J.A., Glöckner,G., Rajandream,M.A., Sucgang,R., Berriman,M., Song,J., Olsen,R., Szafranski,K., Xu,Q. *et al.* (2005) The genome of the social amoeba *Dictyostelium discoideum*. *Nature*, **435**, 43–57.
13. Dubin,M., Fuchs,J., Graf,R., Schubert,I. and Nellen,W. (2010) Dynamics of a novel centromeric histone variant CenH3 reveals the evolutionary ancestral timing of centromere biogenesis. *Nucleic Acids Res.*, **38**, 7526–7537.
14. Glöckner,G. and Heidel,A.J. (2009) Centromere sequence and dynamics in *Dictyostelium discoideum*. *Nucleic Acids Res.*, **37**, 1809–1816.
15. Cappello,J., Cohen,S.M. and Lodish,H.F. (1984) *Dictyostelium* transposable element DIRS-1 preferentially inserts into DIRS-1 sequences. *Mol. Cell. Biol.*, **4**, 2207–2213.
16. Winckler,T., Schiefner,J., Spaller,T. and Siol,O. (2011) *Dictyostelium* transfer RNA gene-targeting retrotransposons: studying mobile element-host interactions in a compact genome. *Mob. Genet. Elements*, **1**, 145–150.
17. Cappello,J., Handelsman,K. and Lodish,H.F. (1985) Sequence of *Dictyostelium* DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. *Cell*, **43**, 105–115.
18. Cohen,S.M., Cappello,J. and Lodish,H.F. (1984) Transcription of *Dictyostelium discoideum* transposable element DIRS-1. *Mol. Cell. Biol.*, **4**, 2332–2340.
19. Rosen,E., Sivertsen,A. and Firtel,R.A. (1983) An unusual transposon encoding heat shock inducible and developmentally regulated transcripts in *Dictyostelium*. *Cell*, **35**, 243–251.
20. Zuker,C., Cappello,J., Chisholm,R.L. and Lodish,H.F. (1983) A repetitive *Dictyostelium* gene family that is induced during differentiation and by heat shock. *Cell*, **34**, 997–1005.
21. Kuhlmann,M., Borisova,B.E., Kaller,M., Larsson,P., Stach,D., Na,J., Eichinger,L., Lyko,F., Ambros,V., Söderbom,F. *et al.* (2005) Silencing of retrotransposons in *Dictyostelium* by DNA methylation and RNAi. *Nucleic Acids Res.*, **33**, 6405–6417.
22. Pall,G.S., Codony-Servat,C., Byrne,J., Ritchie,L. and Hamilton,A. (2007) Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic Acids Res.*, **35**, e60.
23. Southern,E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.*, **98**, 503–517.
24. Hughes,J.E. and Welker,D.L. (1988) A mini-screen technique for analyzing nuclear DNA from a single *Dictyostelium* colony. *Nucleic Acids Res.*, **16**, 2338.
25. Green,M.R. and Sambrook,J. (2012) *Molecular Cloning: a Laboratory Manual*. 4th edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
26. Levi,S., Polyakov,M. and Egelhoff,T.T. (2000) Green fluorescent protein and epitope tag fusion vectors for *Dictyostelium discoideum*. *Plasmid*, **44**, 231–238.
27. Faix,J., Kreppel,L., Shaulsky,G., Schleicher,M. and Kimmel,A.R. (2004) A rapid and efficient method to generate multiple gene disruptions in *Dictyostelium discoideum* using a single selectable marker and the Cre-loxP system. *Nucleic Acids Res.*, **32**, e143.
28. Veltman,D.M., Akar,G., Bosgraaf,L. and Van Haastert,P.J. (2009) A new set of small, extrachromosomal expression vectors for *Dictyostelium discoideum*. *Plasmid*, **61**, 110–118.
29. Hofmann,P., Kruse,J. and Hammann,C. (2013) Transcript localization in *D. discoideum* cells by RNA FISH. *Methods Mol. Biol.*, **983**, 311–323.
30. Bilzer,A., Dolz,H., Reinhardt,A., Schmith,A., Siol,O. and Winckler,T. (2011) The C-module-binding factor supports amplification of TRE5-A retrotransposons in the *Dictyostelium discoideum* genome. *Eukaryot. Cell*, **10**, 81–86.
31. Lucas,J., Bilzer,A., Moll,L., Zundorf,I., Dingermann,T., Eichinger,L., Siol,O. and Winckler,T. (2009) The carboxy-terminal domain of *Dictyostelium* C-module-binding factor is an independent gene regulatory entity. *PLoS One*, **4**, e5012.
32. Pfaffl,M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.*, **29**, e45.
33. Hällman,J., Avesson,L., Reimegård,J., Käller,M. and Söderbom,F. (2013) Identificaton and verification of microRNAs by high-throughput sequencing. *Methods Mol. Biol.*, **983**, 125–138.
34. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.*, **17**, 10–12.
35. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
36. Jurka,J., Kapitonov,V.V., Pavlicek,A., Klonowski,P., Kohany,O. and Walichiewicz,J. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.*, **110**, 462–467.
37. Robinson,J.T., Thorvaldsdottir,H., Winckler,W., Guttman,M., Lander,E.S., Getz,G. and Mesirov,J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.
38. Wiegand,S., Kruse,J., Gronemann,S. and Hammann,C. (2011) Efficient generation of gene knockout plasmids for *Dictyostelium discoideum* using one-step cloning. *Genomics*, **97**, 321–325.
39. Avesson,L., Reimegard,J., Wagner,E.G. and Söderbom,F. (2012) MicroRNAs in Amoebozoa: deep sequencing of the small RNA population in the social amoeba *Dictyostelium discoideum* reveals developmentally regulated microRNAs. *RNA*, **18**, 1771–1782.
40. Werner,A. (2005) Natural antisense transcripts. *RNA Biol.*, **2**, 53–62.
41. Faghihi,M.A. and Wahlestedt,C. (2009) Regulatory roles of natural antisense transcripts. *Nat. Rev. Mol. Cell Biol.*, **10**, 637–643.
42. Aspegren,A., Hinas,A., Larsson,P., Larsson,A. and Soderbom,F. (2004) Novel non-coding RNAs in *Dictyostelium discoideum* and their expression during development. *Nucleic Acids Res.*, **32**, 4646–4656.
43. Hinas,A., Reimegard,J., Wagner,E.G., Nellen,W., Ambros,V.R. and Soderbom,F. (2007) The small RNA repertoire of Dictyostelium discoideum and its regulation by components of the RNAi pathway. *Nucleic Acids Res.*, **35**, 6714–6726.
44. Pak,J. and Fire,A. (2007) Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science*, **315**, 241–244.
45. Klahre,U., Crete,P., Leuenberger,S.A., Iglesias,V.A. and Meins,F. Jr (2002) High molecular weight RNAs and small interfering RNAs induce systemic posttranscriptional gene silencing in plants. *Proc. Natl Acad. Sci. USA*, **99**, 11981–11986.
46. Braunstein,T.H., Moury,B., Johannessen,M. and Albrechtsen,M. (2002) Specific degradation of 3′ regions of GUS mRNA in posttranscriptionally silenced tobacco lines may be related to 5′-3′ spreading of silencing. *RNA*, **8**, 1034–1044.

47. Yang,N. and Kazazian,H.H. Jr (2006) L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat. Struct. Mol. Biol.*, **13**, 763–771.
48. Watanabe,T., Takeda,A., Tsukiyama,T., Mise,K., Okuno,T., Sasaki,H., Minami,N. and Imai,H. (2006) Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev.*, **20**, 1732–1743.
49. Watanabe,T., Totoki,Y., Toyoda,A., Kaneda,M., Kuramochi-Miyagawa,S., Obata,Y., Chiba,H., Kohara,Y., Kono,T., Nakano,T. *et al.* (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, **453**, 539–543.
50. Sijen,T. and Plasterk,R.H. (2003) Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature*, **426**, 310–314.
51. Chung,W.J., Okamura,K., Martin,R. and Lai,E.C. (2008) Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. *Curr. Biol.*, **18**, 795–802.
52. Ghildiyal,M., Seitz,H., Horwich,M.D., Li,C., Du,T., Lee,S., Xu,J., Kittler,E.L., Zapp,M.L., Weng,Z. *et al.* (2008) Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science*, **320**, 1077–1081.
53. Bernstein,E., Caudy,A.A., Hammond,S.M. and Hannon,G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, **409**, 363–366.
54. Ketting,R.F., Fischer,S.E., Bernstein,E., Sijen,T., Hannon,G.J. and Plasterk,R.H. (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.*, **15**, 2654–2659.
55. Knight,S.W. and Bass,B.L. (2001) A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science*, **293**, 2269–2271.
56. Sijen,T., Steiner,F.A., Thijssen,K.L. and Plasterk,R.H. (2007) Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science*, **315**, 244–247.
57. Makeyev,E.V. and Bamford,D.H. (2002) Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Mol. Cell*, **10**, 1417–1427.
58. Moissiard,G., Parizotto,E.A., Himber,C. and Voinnet,O. (2007) Transitivity in *Arabidopsis* can be primed, requires the redundant action of the antiviral Dicer-like 4 and Dicer-like 2, and is compromised by viral-encoded suppressor proteins. *RNA*, **13**, 1268–1278.
59. Curaba,J. and Chen,X. (2008) Biochemical activities of *Arabidopsis* RNA-dependent RNA polymerase 6. *J. Biol. Chem.*, **283**, 3059–3066.
60. Allen,E., Xie,Z., Gustafson,A.M. and Carrington,J.C. (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell*, **121**, 207–221.
61. Schiebel,W., Haas,B., Marinkovic,S., Klanner,A. and Sanger,H.L. (1993) RNA-directed RNA polymerase from tomato leaves. I. Purification and physical properties. *J. Biol. Chem.*, **268**, 11851–11857.
62. Schiebel,W., Haas,B., Marinkovic,S., Klanner,A. and Sanger,H.L. (1993) RNA-directed RNA polymerase from tomato leaves. II. Catalytic *in vitro* properties. *J. Biol. Chem.*, **268**, 11858–11867.

6.4 Manuskript 4

A host factor supports retrotransposition of the TRE5-A population in *Dictyostelium* cells by suppressing an Argonaute protein

---

Anika Schmith, Thomas Spaller, Åsa Fransson, Jonas Kjellin, Benjamin Boesler, Sandeep Ojha, Wolfgang Nellen, Christian Hammann, Fredrik Söderbom, Thomas Winckler

Publiziert in: *Mobile DNA, 2015*

Schmith et al. Mobile DNA (2015) 6:14
DOI 10.1186/s13100-015-0045-5

Mobile DNA

RESEARCH Open Access

A host factor supports retrotransposition of the TRE5-A population in *Dictyostelium* cells by suppressing an Argonaute protein

Anika Schmith[1], Thomas Spaller[1], Friedemann Gaube[1], Åsa Fransson[2,6], Benjamin Boesler[3], Sandeep Ojha[4], Wolfgang Nellen[3,7], Christian Hammann[4], Fredrik Söderbom[5] and Thomas Winckler[1*]

Abstract

Background: In the compact and haploid genome of *Dictyostelium discoideum* control of transposon activity is of particular importance to maintain viability. The non-long terminal repeat retrotransposon TRE5-A amplifies continuously in *D. discoideum* cells even though it produces considerable amounts of minus-strand (antisense) RNA in the presence of an active RNA interference machinery. Removal of the host-encoded C-module-binding factor (CbfA) from *D. discoideum* cells resulted in a more than 90 % reduction of both plus- and minus-strand RNA of TRE5-A and a strong decrease of the retrotransposition activity of the cellular TRE5-A population. Transcriptome analysis revealed an approximately 230-fold overexpression of the gene coding for the Argonaute-like protein AgnC in a CbfA-depleted mutant.

Results: The *D. discoideum* genome contains orthologs of RNA-dependent RNA polymerases, Dicer-like proteins, and Argonaute proteins that are supposed to represent RNA interference pathways. We analyzed available mutants in these genes for altered expression of TRE5-A. We found that the retrotransposon was overexpressed in mutants lacking the Argonaute proteins AgnC and AgnE. Because the *agnC* gene is barely expressed in wild-type cells, probably due to repression by CbfA, we employed a new method of promoter-swapping to overexpress *agnC* in a CbfA-independent manner. In these strains we established an in vivo retrotransposition assay that determines the retrotransposition frequency of the cellular TRE5-A population. We observed that both the TRE5-A steady-state RNA level and retrotransposition rate dropped to less than 10 % of wild-type in the *agnC* overexpressor strains.

Conclusions: The data suggest that TRE5-A amplification is controlled by a distinct pathway of the *Dictyostelium* RNA interference machinery that does not require RNA-dependent RNA polymerases but involves AgnC. This control is at least partially overcome by the activity of CbfA, a factor derived from the retrotransposon's host. This unusual regulation of mobile element activity most likely had a profound effect on genome evolution in *D. discoideum*.

Keywords: *Dictyostelium*, Retrotransposition, siRNA, RNAi, Argonaute

* Correspondence: t.winckler@uni-jena.de
[1]Department of Pharmaceutical Biology, Institute of Pharmacy, University of Jena, Semmelweisstrasse 10, 07743 Jena, Germany
Full list of author information is available at the end of the article

Background

Transposable elements are found in virtually all organisms and play central roles in shaping their host's genomes. The amplification of these genomic parasites is a constant threat to host fitness due to the intrinsic process of integration into the genomic DNA that can cause mutagenesis of genes and force illegitimate recombinations between distant transposon copies [1–4]. Eukaryotic cells have evolved several pathways of RNA interference (RNAi) to restrain the amplification of transposons at the posttranscriptional level [5–8]. In this process, long RNA duplexes (dsRNA), which may occur in cells as intermediates of transposon or RNA virus replication, are typically processed into 20–30 nucleotide double-stranded small interfering RNAs (siRNAs) by ribonuclease III-type enzymes such as Dicer. The siRNAs are loaded onto RNA-induced silencing complexes (RISCs), which are minimally composed of an Argonaute protein and a small RNA [9, 10]. Argonaute proteins are characterized by an RNA-binding PAZ (Piwi-Argonaut-Zwille) domain and a catalytic, ribonuclease H-like PIWI (P-element-induced wimpy testis) domain. Argonaute proteins bind siRNAs via their PAZ domains, unwind the siRNA duplex and use one of the single-stranded RNA molecules as guides to bind mRNAs in a sequence-specific manner [9]. If the guide RNA is fully complementary to the target RNA across the active site of the Argonaute protein, the enzyme is able to degrade the target RNA by a single endonucleolytic cut executed by the PIWI domain, a function termed slicing. If slicing is precluded by mismatches between the annealing guide RNA and cellular mRNA, translation is repressed and mRNA can be degraded by deadenylation and decapping.

The social amoeba *Dictyostelium discoideum* has a haploid genome in which nearly two thirds of DNA are protein-coding genes [11]. Despite the remarkable compactness of its genome, *D. discoideum* accommodates a large number of mobile elements that add up to approximately 10 % of the entire genomic DNA [12]. Most likely for the purpose of suppressing transposition, the organism has evolved a sophisticated RNAi machinery that includes, for example, three RNA-dependent RNA polymerases (RdRPs), two Dicer-like proteins, and five Argonaute-like proteins [13–17]. Intriguingly, the non-long terminal repeat retrotransposon TRE5-A has established a fairly high amplification rate in growing *D. discoideum* cells [18, 19] despite the constitutive production of minus-strand RNA from an element-internal promoter [20, 21]. Thus, how TRE5-A manipulates the cellular RNAi machinery to maintain its remarkable retrotransposition activity is of interest.

Clearly, *D. discoideum* cells could take advantage of TRE5-A's minus-strand RNA production to downregulate TRE5-A plus-strand RNA, the substrate for retrotransposition, using an RNAi pathway. This strategy is actually realized in the silencing of the tyrosine recombinase retrotransposon DIRS-1 in *D. discoideum* cells [22]. To suppress TRE5-A amplification, promoter activity of the C-module, the distinguished minus-strand RNA promoter at the 3′ end of the TRE5-A element, could be positively regulated by a host-encoded transcription factor. This could elevate the level of TRE5-A-derived dsRNA, which could be processed into small RNAs that guide Argonaute proteins to degrade TRE5-A plus-strand RNA and prevent retrotransposition. Consistent with this idea, we previously isolated the C-module-binding factor (CbfA), a host-encoded DNA-binding protein that interacts with the C-module of TRE5-A in vitro [23–25].

The gene CbfA-coding could not be inactivated by conventional homologous recombination (knockout) and may be essential for the growth of *D. discoideum* cells. We constructed a knock-in mutant, JH.D, in which the *cbfA* gene was replaced by a *cbfA* variant containing an *amber* stop codon at amino acid position 455 [25]. The expression of an *amber* suppressor tRNA gene in *D. discoideum* cells allows read-through translation without causing an inherent phenotype [26]. Due to the low efficacy of this *amber* suppression, JH.D cells produce less than 5 % of full-length CbfA protein from the expressed *cbfA*(*amber*) mRNA [25].

JH.D cells have an aberrant developmental phenotype that can be fully rescued by ectopic expression of CbfA in the mutant [27]. Transcriptome analyses revealed that CbfA has general gene regulatory functions in *D. discoideum* cells [28], making this protein an attractive candidate as a host protein that could limit TRE5-A expression and retrotransposition by elevating TRE5-A-derived minus-strand RNA. Interestingly, we observed that both plus- and minus-strand RNA of TRE5-A were reduced concurrently in the CbfA mutant by more than 90 %, and this reduction of transcript levels was accompanied by a sharp drop in TRE5-A's retrotransposition activity in vivo [21]. Remarkably, the promoter activity of neither the A-module (TRE5-A's plus-strand RNA promoter) nor the C-module was altered in reporter gene assays in the CbfA mutant compared to wild-type cells [21]. Thus, we hypothesized that CbfA supports TRE5-A amplification indirectly by down-regulating one or several components of the cellular RNAi machinery. In support of this assumption, a previous transcriptome analysis revealed an approximately 230-fold and 3-fold overexpression of the genes encoding *D. discoideum* Argonaute-like proteins AgnC and AgnE, respectively, in the CbfA-depleted mutant [28].

Here, we found that TRE5-A expression was elevated in knockout strains of *agnC* and *agnE*, suggesting that CbfA may support the accumulation of TRE5-A transcripts by suppressing an RNAi pathway that involves these Argonaute proteins. To determine whether control

of TRE5-A expression by AgnC and/or AgnE leads to a reduction in TRE5-A retrotransposition in vivo, we first developed a new gene activation (GA) strategy to construct strains that overexpress *agnC* in the absence of any residual plasmid sequences inserted in their genomes. We found that the accumulation of TRE5-A RNA was reduced in both $agnC^{GA}$ and $agnE^{GA}$ strains. Next, we employed the previously developed "TRE trap" retrotransposition assay [18, 19] to determine the retrotransposition activity of the cellular TRE5-A population in $agnC^{GA}$ cells. The retrotransposition frequency of the cellular TRE5-A population was determined to be less than 10 % of the wild-type level, suggesting that TRE5-A amplification in *D. discoideum* cells is under surveillance of a distinct RNAi pathway that requires AgnC function and that this control of mobile element expansion is at least in part overcome by CbfA, a factor derived from the retrotransposon's host cell.

Results

CbfA regulates the expression of the Argonaute-like protein AgnC

Even though the accumulation of TRE5-A RNA in *D. discoideum* cells strictly depends on CbfA and this factor binds to the C-module of TRE5-A in vitro, it does not regulate the C-module's promoter activity in vivo [21]. A probable explanation for this paradox could be that CbfA exerts an indirect effect by regulating an RNAi pathway that is involved in the control of TRE5-A expression. In concordance with a rather indirect and probably broader function of CbfA in the control of mobile elements, including TRE5-A, the re-evaluation of previously obtained mRNA-seq data [28] suggested a considerable amount of deregulation of transposable elements in the CbfA-depleted mutant JH.D compared to the parental strain AX2 (Fig. 1). Typically the differential expression of mobile elements between AX2 and JH.D cells was highly variable among different cell cultures and could not be unequivocally verified by quantitative RT-PCR. Nevertheless, this observation strengthened the hypothesis that the regulation of steady-state levels of TRE5-A RNA may not be directly regulated by CbfA by binding to the C-module, but rather be the result of CbfA regulating an RNAi pathway involved in the regulation of TRE5-A.

Following this hypothesis we used previously obtained mRNA-seq data [28] to determine differential expression of putative RNAi components between the CbfA mutant JH.D and parent AX2 cells. The Argonaute genes *agnC* and *agnE* were 228-fold and a 2.7-fold, respectively, overexpressed in JH.D cells (Fig. 2). The genes coding for Argonaute proteins AgnA and AgnB and the RdRP RrpC were slightly underexpressed in the JH.D mutant cells, whereas expression of the genes coding for the

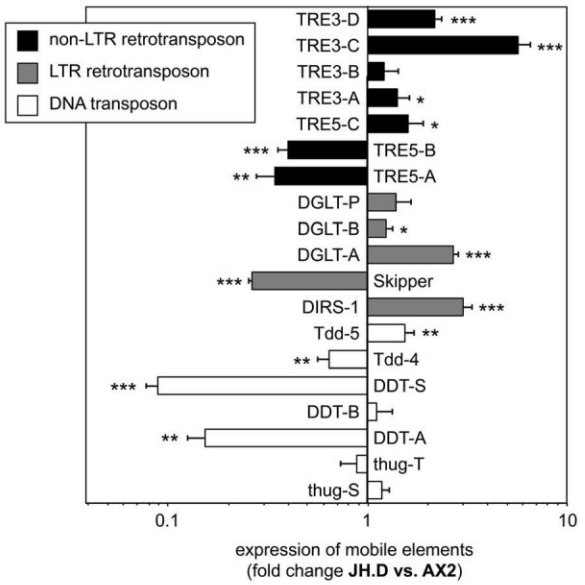


Fig. 1 Expression of mobile elements in the CbfA underexpressing mutant JH.D. Expression data are derived from a previous RNA-seq analysis [28]. RPKM values (reads per kb of mRNA and standardized to 1 million reads) were obtained for the indicated mobile elements and calculated as ratio of JH.D versus AX2 and are shown as "fold change" of expression, meaning that values >1 represent overexpression of genes in JH.D. Values are means from three independent cultures ± SD. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, relative to control AX2 cells (Student's t-test). Black bars indicate non-LTR retrotransposons, grey bars indicate LTR retrotransposons (including the tyrosine recombinase retrotransposon DIRS-1), and white bars indicate putative DNA transposons

RdRPs *rrpA* and *rrpB* were unaffected by CbfA depletion (Fig. 2). RNA-seq also revealed normal expression of the genes coding for the two Dicer-like proteins of *D. discoideum*, *drnA* and *drnB*, in the CbfA mutant (Fig. 2). To confirm the RNA-seq data, we determined the expression the Dicer genes, the three RdRPs, and the five Argonaute genes by qRT-PCR. For these measurements we combined three RNA samples used in the previous RNA-seq experiment with RNA preparations from three additional independent cultures. The data were consistent in all six biological replicates and are presented in Fig. 2. The strong overexpression of *agnC* in the CbfA mutant was confirmed (257-fold, $p < 0.01$, Student's t-test). The weak overexpression of *agnE* seen in RNA-seq could not be verified by qRT-PCR at a statistically significant level (4.3-fold overexpression; $p = 0.17$), although the trend to *agnE* overexpression in JH.D was reproduced (Fig. 2). The weak but highly significant underexpression of *agnA* in the JH.D mutant observed by RNA-seq (2.1-fold; $p < 0.001$) was confirmed by qRT-PCR (3.6-fold; $p < 0.01$), whereas results for *agnB* were inconclusive (Fig. 2).

CbfA can be divided into an amino-terminal part containing a "carboxy-terminal jumonji domain" and two zinc

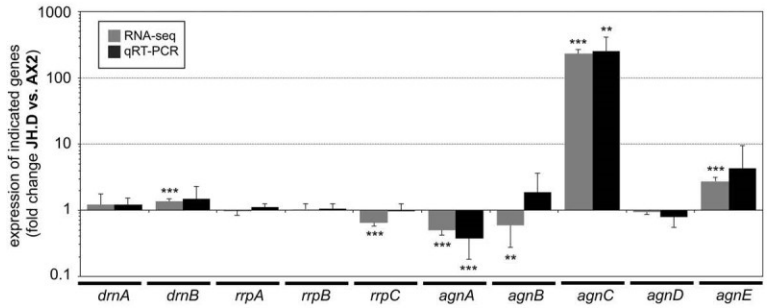


Fig. 2 Expression of RNAi components in CbfA mutant JH.D cells. Expression of Dicer-like proteins (*drnA*, *drnB*), RdRPs (*rrpA-C*) and Argonaute genes (*agnA-E*) was analyzed by RNA-seq (gray bars, n = 3) in wild-type AX2 and CbfA-mutant JH.D cells from three independent cultures [28]. These three RNA samples, and RNA from three additional independent cultures, were analyzed by qRT-PCR (black bars, n = 6). Data are shown as fold change of expression with values >1 meaning that expression in the mutant cells was higher than in the control cells. Values are means ± SD from three and six independent cultures, respectively. **p < 0.01, ***p < 0.001, relative to wild-type AX2 cells (Student's t-test)

finger-like motifs, as well as a carboxy-terminal domain (CbfA-CTD) that contains a DNA-binding AT hook motif (Fig. 3a). We have previously determined that the CbfA-CTD is able to mediate most of CbfA's gene-regulatory activity without requiring the rest of the CbfA protein [28] and it also completely restores TRE5-A expression in JH.D cells [21]. We wanted to evaluate whether the aberrant expression of *agnC* and *agnE* in JH.D cells was reversed by expression of CbfA in the mutant. To this end, we expressed either full-length CbfA or a GFP fusion to CbfA-CTD in JH.D cells using multicopy expression plasmids as shown in Fig. 3b. To confirm functional expression of CbfA or CbfA-CTD in JH.D, we first determined TRE5-A expression (Additional file 1: Figure S1). TRE5-A underexpression was rescued to wild-type level by full-length CbfA and was even "over-complemented" by CbfA-CTD, which was likely due to overexpression of this protein relative to normal CbfA amounts present in AX2 cells (compare Fig. 3b). The overexpression of *agnC* in JH.D was reduced by full-length CbfA by 80 % ($p < 0.05$) and by CbfA-CTD by 92 % ($p < 0.05$) (Fig. 3c). This result was similar to previous RNA-seq data [28], which revealed 97 % reduction ($p < 0.001$) of *agnC* overexpression in response to the presence of CbfA-CTD [28]. Taken together, the data indicate that *agnC* is a genuine CbfA-regulated gene that requires CbfA-CTD for proper expression. The data argued for a role of AgnC in CbfA-dependent TRE5-A regulation because the accumulation of TRE5-A transcripts is also regulated by CbfA-CTD [21].

Expression of full-length CbfA in JH.D cells had only a minor effect on the observed overexpression of *agnE* in JH.D cells. Likewise, expression of CbfA-CTD in JH.D cells did not affect *agnE* overexpression in the CbfA mutant (Additional file 1: Figure S1). The latter results were consistent with previous RNA-seq data, which did not indicate an effect of CbfA-CTD on *agnE* expression [28]. Therefore, we cannot definitely conclude from our data that *agnE* is a genuine CbfA-regulated gene.

AgnC and AgnE downregulate TRE5-A expression

We performed qRT-PCR measurements to determine whether putative components of the *D. discoideum* RNAi machinery are involved in the silencing of TRE5-A expression. No significant changes of TRE5-A expression were determined when the genes coding for the Dicer-like protein DrnB or the RdRP proteins RrpA and RrpB were inactivated; however, we detected a mild but significant underexpression of TRE5-A in *rrpC* knockout cells (Fig. 4). Inactivation of *agnA* or *agnB* had no effect on TRE5-A expression (Fig. 4), whereas a 4.3- and 5.9-fold overexpression of TRE5-A was observed in knockout mutants of *agnC* and *agnE*, respectively (Fig. 4). Overexpression of TRE5-A in *agnC* and *agnE* knockout strains was completely reversed when AgnC or AgnE was expressed from a multicopy plasmid in the respective knockout mutant (Fig. 4), suggesting that both AgnC and AgnE contribute to TRE5-A regulation.

The role of AgnC and AgnE in the suppression of the TRE5-A population could be analyzed in more detail if the corresponding genes would be overexpressed in a wild-type background, i.e., in a strain with normal CbfA activity. This was an important consideration because previous data indicated that CbfA may have functions in TRE5-A retrotransposition beyond the regulation of TRE5-A RNA levels in supporting the integration process upstream of tRNA genes [21]. Usually, overexpression of proteins in wild-type cells is facilitated by transforming cells with expression plasmids. We assumed that this would be a suboptimal strategy for our experiments because transformants would contain insertions of multicopy plasmids at random genomic positions that could compromise the subsequent determination of retrotransposition activity of the TRE5-A population using the TRE trap assay (see below). With this consideration in mind, we decided to generate gene activation (GA) strains in which the endogenous promoter of either *agnC* or *agnE* was replaced by the strong *actin15* promoter (*act15P*) by

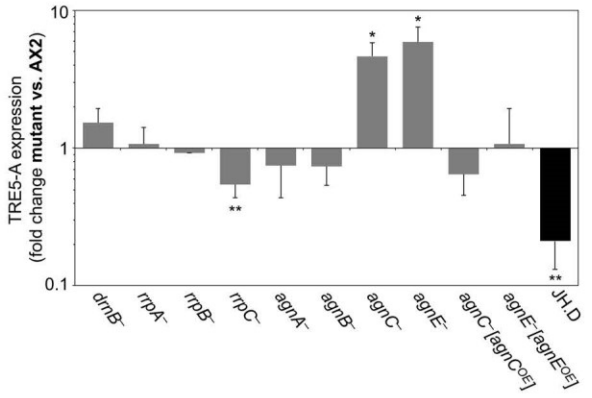


Fig. 4 Expression of TRE5-A in knockout mutants of RNAi components. Expression of TRE5-A ORF1 was analyzed by qRT-PCR in the indicated knockout mutants of Argonaute genes and the Dicer-like protein DrnB. Phenotype reversion in *agnC* and *agnE* knockouts was accomplished using TAP-tagged *agnC* and *agnE* overexpressed in the respective mutants (*agnC* [$agnC^{OE}$] and *agnE* [$agnE^{OE}$]). TRE5-A expression in JH.D cells is shown for comparison. Expression levels were compared to AX2 wild-type cells and are expressed as fold change of expression, meaning that values >1 represent overexpression of genes in the mutants. Data represent means from three independent cultures ± SD. *$p < 0.05$, **$p < 0.01$ relative to AX2 cells (Student's t-test)

2016-fold excess *agnC* mRNA compared to AX2 cells. Yet the $agnC^{GA}$ mutants had no obvious phenotype during growth and multicellular development. Expression of the *agnC*-upstream gene DDB_G0271884 was not affected by the homologous recombination yielding the $agnC^{GA}$ strains (Fig. 5b). Semi-quantitative RT-PCR revealed that the TRE5-A steady-state transcript level in the $agnC^{GA}$ mutants dropped sharply and was even more lower than in JH.D cells (Fig. 5b). This was confirmed by qRT-PCR, which suggested 24- to 45-fold lower expression of the TRE5-A ORF1 transcript in $agnC^{GA}$ mutants than in wild-type cells, compared to a 4-fold decrease seen in JH.D cells (Fig. 5c). The data suggested that AgnC is directly involved in suppressing TRE5-A transcripts in *D. discoideum* cells.

Strains overexpressing the *agnE* gene in the AX2 background were constructed by employing the gene activation strategy as outlined in Additional file 1: Figure S2. Apparently overexpression of AgnE in the recovered $agnE^{GA}$ strains resulted in an average of 2.3-fold underexpression of TRE5-A in four strains tested, thus confirming that AgnE may be involved in TRE5-A suppression.

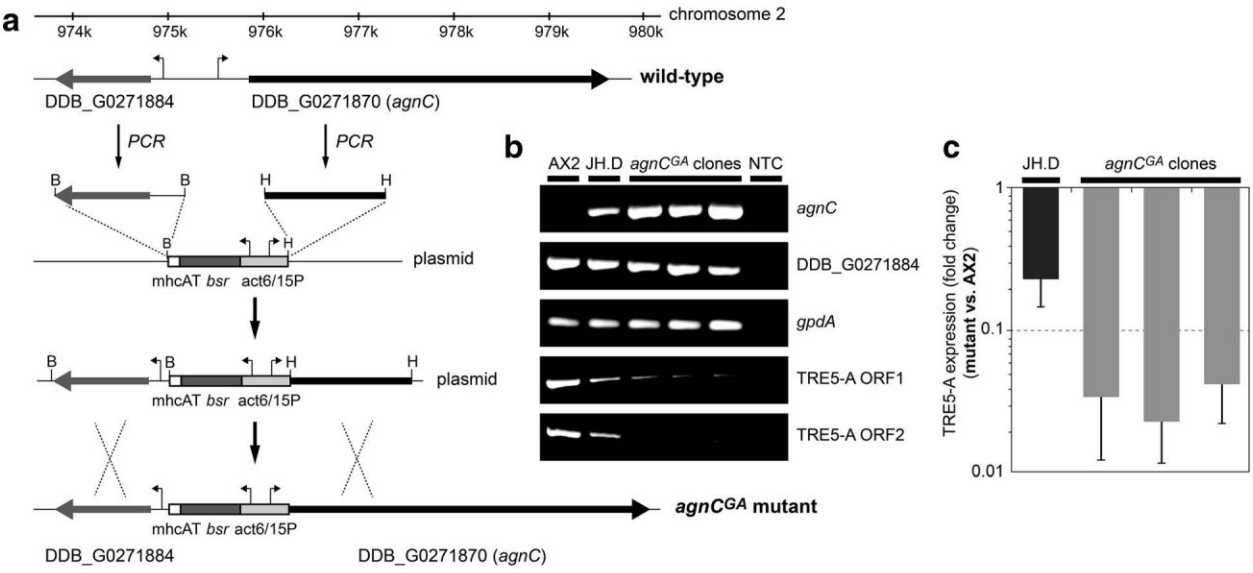


Fig. 5 TRE5-A expression in $agnC^{GA}$ mutants. **a** Construction of *agnC* "gene activation" mutants. The *agnC* locus on chromosome 2 is indicated by nucleotide positions. The gene activation cassette consisted of a hybrid *actin6/actin15* promoter (arrows indicate transcription direction). The BamHI arm contained a 1200 bp DNA fragment covering the complete coding sequence of gene DDB_G0271884, including 273 bp of upstream sequence. The HindIII arm contained 1200 bp of *agnC* coding sequence, including the original translation start site. After double-recombination of the $agnC^{GA}$ vector with genomic DNA, the expression of *agnC* was driven by the *act15* promoter, whereas expression of the neighboring gene DDB_G0271884 was unaffected. **b** Semi-quantitative RT-PCR analysis of RNA from AX2, JH.D, and three independent $agnC^{GA}$ mutants demonstrating overexpression of *agnC*, normal expression of the neighboring gene DDB_G0271884 and *gpdA* (loading control), and silencing of TRE5-A (ORF1 and ORF2 sequences). NTC: no template control. **c** Quantitative RT-PCR of TRE5-A (ORF1) expression on RNA from JH.D and three $agnC^{GA}$ mutants. Expression levels were compared to AX2 cells and are expressed as fold change of expression, meaning that values <1 represent lower levels of TRE5-A in the mutants relative to wild-type AX2 cells. Data represent means from four independent cultures of the indicated strains ± SD

AgnC is a suppressor of TRE5-A retrotransposition

It seemed plausible that silencing of TRE5-A expression would diminish retrotransposition of the TRE5-A population in the *agnC* and *agnE* overexpressor strains. To directly measure the retrotransposition activity of the cellular TRE5-A population, we set up a previously described "TRE trap" in vivo retrotransposition assay [18, 19] in $agnC^{GA}$ cells. The TRE trap is based on a modified *pyr56* gene (TRE^{trap}) that codes for UMP synthase (Fig. 6). The TRE^{trap} gene contains an intron into which a Val^{UAC} tRNA gene was inserted as target for TRE5-A integrations. After transformation of the TRE^{trap} gene into ura$^-$ cells, the transformants present a ura$^+$ phenotype due to the expression of functional UMP synthase from the TRE^{trap} gene (with the intron including the tRNA gene being spliced out); however, the cells are prone to mutations in the TRE^{trap} gene by integration of cellular TRE5-A elements upstream of the embedded tRNA gene. Cells affected by TRE5-A integration into the TRE^{trap} gene can no longer splice out the intron and are converted to the ura$^-$ phenotype; they can be recovered after clonal growth in medium complemented with 5-fluoroorotic acid (5-FOA) and uracil. In previous studies, approximately 100 insertions into the TRE trap were analyzed for integration by mobile elements [18, 19]. The data revealed that ~1 % of recovered ura$^-$ clones had spontaneous loss-of-function mutations of the TRE^{trap} gene, whereas ~99 % of the clones carried a TRE5-A element. No insertions of other members of the tRNA gene-specific TRE retrotransposon family were detected in this assay, suggesting that they amplify at a very low rate. Thus, the number of clones obtained in the TRE trap assay is an estimate of the TRE5-A retrotransposition activity in *D. discoideum* cells [18, 19].

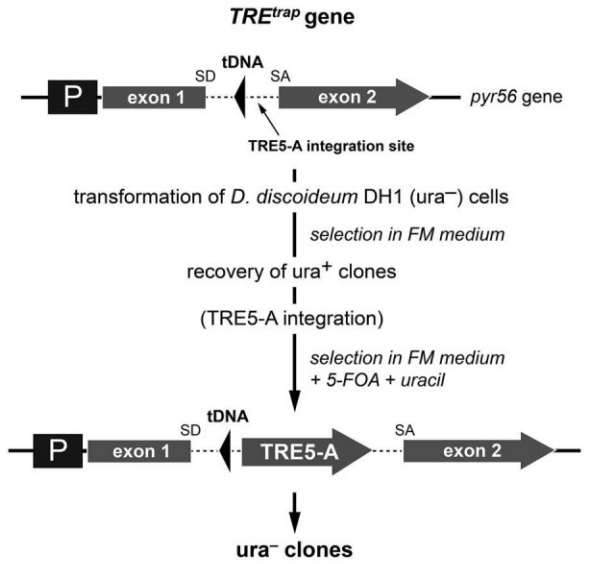


Fig. 6 Outline of the TRE trap retrotransposition assay. The TRE^{trap} gene is a modified version of the *pyr56* gene, which codes for the *D. discoideum* UMP synthase. The *pyr56* gene contains an intron (dashed line; SD: splice donor site, SA: splice acceptor site) into shich a tRNA gene is inserted as bait for the integration of TRE5-A. The TRE^{trap} gene is transformed into the ura$^-$ strain DH1 that has a complete deletion of the *pyr56* gene. The TRE^{trap} gene converts transformants to ura$^+$ because the intron is functionally spliced. If an element of the endogenous TRE5-A population targets the tRNA gene in the TRE^{trap} gene for integration, the TRE^{trap} gene is disrupted even if the integration actually occurs in the intronic sequence. The resulting ura$^-$ cells gain resistance to the drug 5-fluoroorotic acid (5-FOA) and grow out clonally if uracil is added to the medium [18]

In the experiments described above, AX2 was the parent strain for the generation of $agnC^{GA}$ and $agnE^{GA}$ mutants because we wanted to be able to directly compare them to JH.D cells, which were also derived from AX2. Because no suitable uracil-auxotrophic AX2 mutant was available, we reproduced $agnC^{GA}$ mutants in the ura$^-$ strain DH1, which is an AX3 derivative [30]. As shown in Fig. 7a, overexpression of *agnC* in the recovered DH1[$agnC^{GA}$] strains was comparable to AX2[$agnC^{GA}$] cells. Likewise, TRE5-A ORF1 expression was suppressed in DH1[$agnC^{GA}$] strains, albeit at somewhat lower efficacy as in AX2[$agnC^{GA}$] cells. qRT-PCR revealed that TRE5-A expression in the particular DH1[$agnC^{GA}$] strain that we subsequently used to determine TRE5-A retrotransposition was 7.7-fold lower than in the parental DH1 strain. This DH1[$agnC^{GA}$] strain was transformed with plasmids carrying either the empty TRE trap (i.e., no tRNA gene inserted in the trap) or the TRE^{trap} gene, which contained a Val^{UAC} tRNA gene as bait for TRE5-A integrations. Five plates, each containing 10^7 cells, were cultured in minimal medium supplemented with 5-FOA and uracil until clones appeared. As the positive control, the TRE5-A retrotransposition frequency in DH1[TRE^{trap}] cells was determined at 2.03×10^{-5}, whereas it was $<0.01 \times 10^{-5}$ in DH1[TRE^{trap}] cells in which the Val^{UAC} tRNA gene was omitted as the negative control. In two independently recovered DH1[$agnC^{GA}/TRE^{trap}$] strains, TRE5-A retrotransposition activity was determined at 0.14×10^{-5} and 0.05×10^{-5}, representing a more than 90 % drop retrotransposition in the *agnC* overexpressing cells compared to control cells ($p < 0.001$, Student's t-test) (Fig. 7b). These data indicate that AgnC controls the amplification of TRE5-A elements in *D. discoideum* cells by limiting the accumulation of retrotransposon-derived RNA.

Although we were able to establish $agnE^{GA}$ strains in the DH1 background, we could not recover viable cells after transformation of the TRE^{trap} gene into these cells. Therefore, we were unable to determine the retrotransposition activity of the TRE5-A population under these conditions. Thus, it remains elusive at this point whether the moderate downregulation of

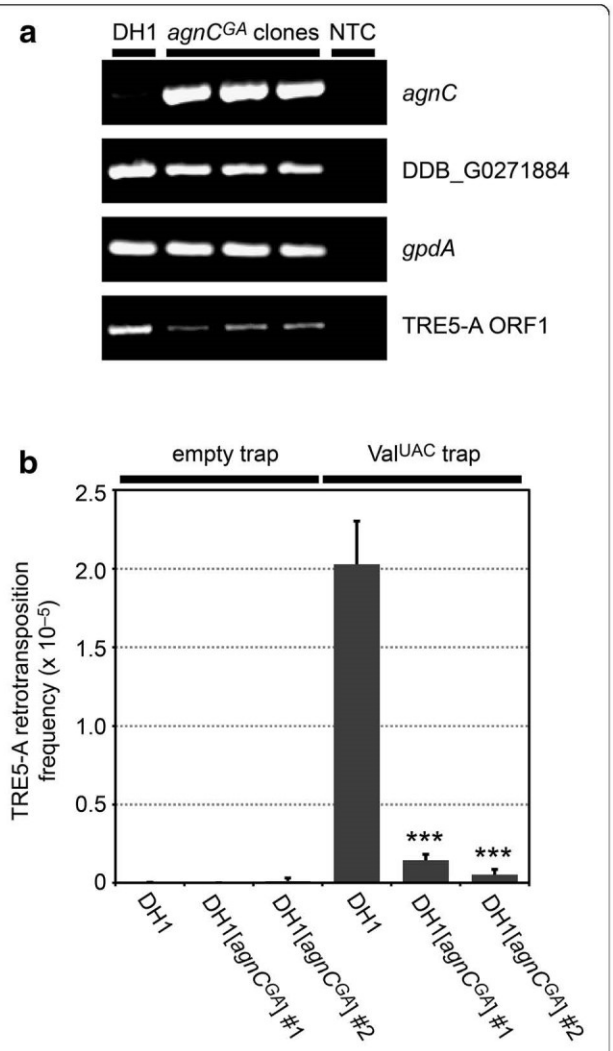


Fig. 7 Retrotransposition of the TRE5-A population in *agnC* overexpressing strains. **a** Semi-quantitative RT-PCR analysis of RNA from DH1 and three independent DH1[$agnC^{GA}$] mutants. Note that *agnC* was overexpressed comapred to DH1. Expression of the neighboring gene DDB_G0271884 (compare Figure 5) was not affected, whereas expression of TRE5-A ORF1 was reduced in the $agnC^{GA}$ mutants. Expression of the house-keeping gene *gpdA* is shown as control. NTC: no template control. **b** DH1 and DH1[$agnC^{GA}$] cells were transformed with either the empty TRE^{trap} gene (i.e., no tRNA gene inserted in the intron) or the TRE^{trap} gene containing a Val^{UAC} tRNA gene. In this TRE trap assay, one clone of DH1 transformants and two independent clones of DH1[$agnC^{GA}$] cells (clones #1 and #2) were analyzed for TRE5-A retrotransposition. For each strain, five plates containing 10^7 cells each were prepared. Cells were cultured in FM medium containing 5-FOA and uracil until clones appeared. Clones were counted and results are presented as retrotransposition frequency ± SD. ***$p < 0.001$ (Student's t-test). This assay was reproduced once with similar results

TRE5-A transcripts in $agnE^{GA}$ strains correlates with an appreciable suppression of the retrotransposition activity of the TRE5-A population.

Discussion

AgnC and AgnE act in RNAi pathways to suppress TRE5-A retrotransposition

Because the TRE5-A element produces both plus- and minus-strand RNA, we assumed that the retrotransposition frequency of the TRE5-A population may be under surveillance by the cellular RNAi machinery. In this study, we provide evidence to support this assumption. Genetic inactivation of either AgnC or AgnE resulted in overexpression of TRE5-A, suggesting that both proteins have functions in TRE5-A regulation in which the loss of one cannot be compensated for by expression of the other. The *D. discoideum* Argonaute proteins are a part of the PIWI subfamily of Argonaute proteins. They all have divergent amino-terminal domains, but possess conserved PAZ and PIWI domains including an intact DEDH catalytic tetrad and probably possess slicer activity. One model to explain TRE5-A silencing is the generation of PIWI-interacting RNAs (piRNAs) by AgnC and AgnE in a ping-pong piRNA replication mechanism typical for PIWI proteins [31]. This model is intuitive given that piRNAs are often generated from long single-stranded RNA precursors produced from transposable elements, such as the minus strand RNA of TRE5-A. However, piRNA are usually 23–30 nt long [31, 32], which contradicts results from a previous deep sequencing of small RNA libraries which revealed the formation of ~21 nt siRNAs from TRE5-A elements at a very low level (0.05 % of total small RNAs) in growing *D. discoideum* cells [15].

The silencing of the retrotransposon DIRS-1 is a model to study RNAi pathways in *D. discoideum* [14, 15, 17, 22]. Previous deep sequencing of small RNAs revealed high levels of DIRS-1-derived ~21 nt siRNAs [15] that add up to 20 % of all small RNAs detected in *D. discoideum* cells [17]. The difference in the amount of ~21 nt siRNAs derived from DIRS-1 and TRE5-A may be interpreted as the less efficient dsRNA formation from TRE5-A RNA compared to DIRS-1 RNA. DIRS-1 silencing is enhanced by the RdRP RrpC that synthesizes new DIRS-1 dsRNA that can be diced into secondary siRNAs [16]. This amplification step may be missing in the RNAi pathway that controls TRE5-A expression, because TRE5-A transcripts were not stabilized in RdRP-deficient mutants. The Dicer homolog DrnB, which is mainly required for miRNA formation in *D. discoideum* [15], is apparently not involved in the regulation of DIRS-1 RNA levels [17] and it seems to be also dispensable in the process of TRE5-A regulation (Fig. 4). Thus, the RNAi pathways that regulate DIRS-1 and TRE5-A may overlap at the stage of primary siRNA formation, presumably involving the Dicer homolog DrnA, but use different RISCs that contain either AgnA for DIRS-1 silencing [22] or AgnC/AgnE for TRE5-A suppression.

CbfA abrogates TRE5-A suppression by repressing AgnC

The C-module at the 3′ end of the TRE5-A element has promoter activity that is responsible for the production of minus-strand RNA by the element [20]. CbfA was originally identified as a "C-module-binding factor" because it binds to the C-module of TRE5-A in vitro [24], but it does not regulate the C-module promoter activity in vivo [21]. Considering that CbfA regulates more than 1000 genes of the *D. discoideum* genome, the present data suggest that the observations of in vitro binding of CbfA to the C-module and in vivo regulation of steady-state levels of TRE5-A transcripts by CbfA are purely coincidental. Together with the data obtained in this study, we propose instead that the accumulation of TRE5-A transcripts in *D. discoideum* cells is indirect and a result of the CbfA-mediated suppression of a post-transcriptional pathway involving AgnC. This assumption is supported by the observation that both the plus- and minus-strand RNA of TRE5-A simultaneously vanish upon removal of CbfA from cells, but reappear when CbfA is re-introduced into CbfA-underexpressing cells [21]. In a previous mRNA-seq experiment comparing gene expression in JH.D with wild-type cells [28] we detected underexpression of putative DNA transposons such as DDT-A and DDT-S, the long-terminal repeat (LTR) retrotransposon Skipper, and the non-LTR retrotransposon TRE5-B (see Fig. 1). RNA-seq also predicted overexpression of some mobile elements in the absence of CbfA such as the tyrosine recombinase retrotransposon DIRS-1, the LTR retrotransposon DGLT-A, and the non-LTR retrotransposons TRE3-C and TRE3-D. However, differential expression of the mentioned mobile elements in JH.D cells could not be unequivocally confirmed by qRT-PCR. This was obviously due to high biological variation between independent cultures that was never observed when analyzing the expression of coding genes (i.e., Fig. 2) and the reason for this phenomenon remains elusive. At least the overexpression of DIRS-1 in JH.D cells could be explained by the weak, but reproducible overexpression of the genes *rrpC* and *agnA* (Fig. 2), which were both shown to be involved in the downregulation of DIRS-1 [22]. Thus, DIRS-1 and TRE5-A may be suppressed by different RNAi pathways and are affected indirectly by CbfA's broad-ranging gene-regulatory activity. Unfortunately, a high variability of mobile element expression among biological replicates was also observed when analyzing either *agnC* knockout or *agnC* overexpressor cells. Therefore, we are unable at this point to predict whether other mobile elements are regulated by the same AgnC-involving RNAi pathway that controls TRE5-A retrotransposition.

TRE5-A belongs to a family comprising seven tRNA gene-targeting retrotransposons in *D. discoideum* cells. The TRE5-A and TRE5-B elements are closely related and share a common ancestor, but only TRE5-A was amplified to a high copy number rather late in the evolution of this species [12]. It is puzzling that TRE5-A amplification in *D. discoideum* was apparently accelerated after the acquisition of the C-module (i.e., an antisense promoter) and after the split from its common ancestor with the TRE5-B element that lacks a C-module. Intuitively, the incorporation of an antisense promoter into a mobile element should make it vulnerable to silencing by RNAi-related mechanisms and thus prevent its amplification. Because the C-module was most likely acquired by 3′-transduction, a process not uncommon in this class of retrotransposons [33], the question of what advantage the element may have gained by incorporating an antisense promoter at its 3′ end remains. Did *D. discoideum* cells gain a selective advantage from TRE5-A expansion? The release of TRE5-A from RNAi surveillance by a regulated process involving a host-encoded factor such as CbfA may have evolved because it could be used for cellular purposes such as enhancing genome flexibility. Alternatively, TRE5-A release from suppression by RNAi may have been incidentally caused by adaptation to evolutionary pressure forcing alterations in AgnC-mediated posttranscriptional regulation that are unrelated to transposon suppression. It is unknown under which conditions the repression of *agnC* by CbfA would be released or for which functions AgnC would be required; at least, it seems to be unrelated to the multicellular development of *D. discoideum* because *agnC* is barely upregulated during development [34] and neither *agnC* knockouts nor *agnC* overexpressor display a developmental phenotype. Repression of AgnC by CbfA may provide an efficient way to respond to changes in particular environmental conditions that require specialized functions of this Argonaute protein. Even if this mode of gene regulation by CbfA would come at the cost of TRE5-A amplification, it is reasonably tolerable because TRE5-A's targeted integration to regions upstream of tRNA genes would largely prevent insertion mutagenesis of the genome.

Whereas RNAi may have been developed to restrict mobile element expansion in *D. discoideum* as in other eukaryotes, as exemplified by DIRS-1 silencing, our study shows an intriguing example of a transposable element that is under surveillance by the cellular RNAi machinery, but the control of which can be overcome by suppression of a distinct RNAi pathway by a host factor.

Conclusions

The social amoeba *D. discoideum* has a compact and haploid genome that requires tight control of mobile element activity to maintain genome stability. The non-long terminal repeat retrotransposon TRE5-A actively amplifies in the genome of *D. discoideum* even though

the element should be vulnerable to posttranscriptional silencing due to the production of antisense RNA from an element-internal promoter. The host-encoded factor CbfA has global gene-regulatory functions in *D. discoideum* that include the suppression of the Argonaute-like proteins AgnC and AgnE. Whereas TRE5-A transcipts were found to accumulate in mutants lacking AgnC or AgnE, expression and retrotransposition of the element vanished in AgnC and AgnE overexpressing cells. These observations suggest that TRE5-A amplification is under surveillance by an RNAi pathway that involves AgnC and AgnE and that this control is at least partially overcome by the activity of CbfA. This unusual regulation of mobile element activity by a host factor most likely had a profound effect on genome evolution in *D. discoideum*.

Methods

Strains and plasmids

The CbfA-depleted mutant JH.D and the plasmids used for the expression of full-length CbfA and the GFP-tagged carboxy-terminal domain of CbfA (CbfA-CTD) have been previously described [21, 25]. *D. discoideum* strains harboring knockouts of RdRP genes *rrpA*, *rrpB*, and *rrpC* were described by Wiegand & Hammann [16]. Knockout strains of *agnA* and *agnB* were described elsewhere [22]. The *drnB*$^{-}$ strain was described in Avesson et al., 2012 [35]. Knockout mutants of *agnC* and *agnE* as well as plasmids allowing for the expression of TAP-tagged AgnC and AgnE will be described in a separate publication (F.S. et al., manuscript in preparation).

Construction of gene activation mutants

The *D. discoideum* expression vector pDM326 [36] contains a blasticidin resistance cassette driven by the *act6* promoter and an upstream *act15* promoter in opposite direction for the expression of transgenes. A DNA fragment containing both the blasticidin cassette and the *act15* promoter was isolated from pDM326 by digestion with BamHI and BglII. The DNA fragment was inserted into the BamHI site of pGEM7Zf(−) (Promega), such that the former BglII site was placed next to the HindIII site of the pGEM vector to generate pGEM-GA. To generate the *agnC*^{GA} vector, the "BamHI arm" covering the entire coding sequence of the gene DDB_G0271884, which shares its upstream region with *agnC*, was amplified including 273 bp of residual upstream sequence and inserted into the BamHI site of pGEM-GA. The "HindIII arm" was generated by amplification of nucleotides 1–1166 of the *agnC* gene including its authentic translation start codon. The pGEM-agnC-GA plasmid was linearized and transformed into *D. discoideum* AX2 or DH1 cells and transformants were selected in HL5 medium (Formedium, Hunstanton, UK) containing 6 μg/ml blasticidin (Life Technologies, Carlsbad, USA) [37]. From such clones genomic DNA was isolated and screened by PCR for insertion of the GA cassette at the targeted locus using one primer specific for the blasticidin resistance gene and a second primer that hybridized outside of the DNA sequences covered by the HindIII arm. RT-PCR was used to confirm that the expression of the *agnC*-upstream gene DDB_G0271884 was not affected by insertion of the GA cassette.

Reverse transcription-PCR

Total RNA was prepared from frozen cell pellets and RT-PCR was done as described previously [21]. In quantitative RT-PCR gene expression levels were standardized to the gene coding for catalase (*catA*). The following qRT-PCR primers were used: rrpA-01, GAA CGTCAAGAACTTGGTAAATTGTATC; rrpA-02, TA ACCTACAGTTTGTAAC CGAATGTTTAC; rrpB-01, GAACGTCAAGAACTTGGTAAAATGTATAA; rrpB-02, GTGGATAACCTTTAGTTTTTAACCAAAC; rrpC-01, GGTGTTTATAGTAAAAAAGAATCATTC; rrpC-02, CAACTATCCAAGAATTTATGAACATTTAC; agnA-01, GCCGAAACTCCTTCTTCTTGGGGTAC; agnA-02, GT TCATCCAATAAGACATGGTAATGAG; agnB-01, GTG ATGGTGTTGGTGATGGTATGTTAG; agnB-02, CTTG GTAATCCTGATCAAGGTGTTGTTG; agnC-03, GTG CACTTTTATGAGAGTATTGGCATAC; agnC-04, GT ACATGATAATGAGTTGGATTTGTAG; agnD-01, CA TCATATTAATAGTCGTTTACCAGAG; agnD-02, GT ACCAATCCACCCAATGGTACAATGG; agnE-03, GA GCATAATTACAAGGAGCAGGTGTTC; agnE-04, CA GTGCTAACCATTGTACCATTGGGTG; catA-01, GT TTCGCTGCTCGTCAACCATACAATC; catA-02, GC ACGAACTTGAATTTCTTTGATGGTG; gpdA-01, GG TTGTCCCAATTGGTATTAATGG; gpdA-02, CCGTG GGTTGAATCATATTTGAAC; TRE5-A ORF1 Rep-108, GTCATAAACATCAATCCGAACCAGAC; TRE5-A ORF1 Rep-109, GTTAGATTGTCTAGTTCAATGATAGTGTC; TRE5-A ORF2 Rep-75, GACTGTTCAGTGGATAATAA CC; TRE5-A ORF2 Rep-176, CTCGAGTTAAAGGAAGA TTGCTCTTGAATC; DDB_G0271884-01, GAGTTGGC CAAATTAGTTAAGCAATTG; DDB_G0271884-02, CC TTGTTCAACCCAAGAGAAAATTTCTG.

TRE trap retrotransposition assay

The TRE trap is an in vivo retrotransposition assay that measures the activity of the cellular TRE5-A population. It was essentially performed as described previously [21]. The TRE trap consists of the complete *pyr5-6* gene modified to contain a functional intron into which a *Val*^{UAC} bait tRNA gene was inserted. This gene is referred to as the *TRE*^{trap} gene. After transformation into *D. discoideum* DH1 cells, ura^{+} cells harboring chromosomal integrations of the *TRE*^{trap} gene were recovered by cultivation in FM medium without supplements. After integration of a

TRE5-A element into the trap, the TRE^{trap} gene is disrupted and no functional UMP synthase is expressed. Thus, affected cells were converted to the ura$^-$ phenotype and gained resistance to the drug 5-fluoroorotic acid (5-FOA). In a typical retrotransposition assay, 5 plates each containing 10^7 cells were prepared, and cells were cultured in FM medium containing 150 μg/ml 5-FOA and 20 μg/ml uracil. Clones that arose were counted, and the data presented are the means from 5 plates ± SD.

Western blots

D. discoideum cells were washed in phosphate buffer and stored as frozen pellets of 2×10^7 cells at −80 °C. SDS/polyacrylamide gel electrophoresis of whole-cell extract proteins and western blotting were done as described [38]. We used monoclonal antibody 7 F3 to detect CbfA and a polyclonal antiserum to detect actin8 [38].

Additional data files

The following additional data are available with the online version of this paper. Additional file 1: Figure S1 shows the functional complementation of strain JH.D with CbfA or its carboxy-terminal domain with respect to TRE5-A and *agnE* expression. Additional file 1: Figure S2 illustrates the construction of $agnE^{GA}$ mutants and shows the expression of TRE5-A in these mutants.

Additional file

Additional file 1: Figure S1. Functional complementation of strain JH.D with CbfA or its isolated carboxy-terminal domain. RNA levels of retrotransposon TRE5-A (A) and *agnE* (B) were determined by qRT-PCR in JH.D cells, JH.D cells expressing full-length CbfA, and JH.D cells expressing CbfA-CTD. Expression levels in JH.D cells and JH.D transformants were compared to AX2 wild-type cells and are expressed as "fold change" of expression, meaning that values >1 represent overexpression of genes in the JH.D strains and a value of 1 would indicate complete reversion of the overexpression in JH.D cells. Values are means from six independent cultures ± SD. **$p < 0.01$, relative to control AX2 cells (Student's t-test). Note that TRE5-A is actually overexpressed in the JH.D[CbfA-CTD] transformant, which is an effect of overexpression of CbfA-CTD (see Fig. 2, main text). **Figure S2.** TRE5-A expression in $agnE^{GA}$ mutants. (A) Construction of *agnE* "gene activation" mutants. The *agnE* locus on chromosome 5 is indicated by nucleotide positions. The gene activation cassette consisted of a hybrid *actin6/actin15* promoter (arrows indicate transcription direction). The BamHI arm contained a 1070 bp DNA fragment covering part of the coding sequence of gene DDB_G0289385. The HindIII arm contained 1080 bp of *agnE* coding sequence, including the original translation start site. After double-recombination of the $agnE^{GA}$ vector with genomic DNA, the expression of *agnE* was driven by the *act15* promoter, whereas expression of the neighboring gene DDB_G0289385 was unaffected. (B) Semi-quantitative RT-PCR analysis of RNA from AX2, JH.D, and three independent $agnE^{GA}$ mutants demonstrating overexpression of *agnE*, normal expression of the neighboring gene DDB_G0289385 and *gpdA* (loading control), and silencing of TRE5-A (ORF1 and ORF2 sequences). NTC: no template control. (C) Quantitative RT-PCR of TRE5-A (ORF1) expression on RNA from JH.D and four $agnE^{GA}$ mutants. Expression levels were compared to AX2 cells and are expressed as fold change of expression, meaning that values <1 represent lower levels of TRE5-A in the mutants relative to wild-type AX2 cells. Data represent means from four independent cultures of the indicated strains ± SD. (PDF 1033 kb)

Abbreviations

CbfA: C-module-binding factor; CTD: carboxy-terminal domain (of CbfA); GA: gene activation (by promoter swapping); LTR: long terminal repeat; qRT-PCR: quantitative reverse-transcription PCR; RdRP: RNA-dependent RNA polymerase; RISC: RNA-induced silencing complex; RNAi: RNA interference; siRNA: small interfering RNA; TAP: tandem affinity purification; TRE5-A: tRNA gene-targeted retroelement 5-A.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TW conceived the study. AS, TS, FG, SO, ÅF, and BB designed and performed the experiments and revised the manuscript critically for important intellectual content. TW, WN, FS, and CH analyzed the data and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgments

This work was supported by grants to TW and HC (WI 1142/7-1, HA3459/7-1) from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) and by support from the Swedish Research Council (to FS and to Uppsala RNA Research Center) and The Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS) to FS. SO acknowledges a stipend by the Fritz Thyssen Foundation.

Author details

[1]Department of Pharmaceutical Biology, Institute of Pharmacy, University of Jena, Semmelweisstrasse 10, 07743 Jena, Germany. [2]Department of Molecular Biology, Biomedical Center, Swedish University of Agricultural Sciences, Uppsala, Sweden. [3]Institute of Biology – Genetics, University of Kassel, Kassel, Germany. [4]Ribogenetics@Biochemistry Lab, Department of Life Sciences and Chemistry, Molecular Life Sciences Research Center, Jacobs University Bremen, Bremen, Germany. [5]Department of Cell and Molecular Biology, Biomedical Center, Uppsala University, Uppsala, Sweden. [6]Present address: Aprea AB, Karolinska Institutet Science Park, Nobels väg 3, 17175 Solna, Sweden. [7]Present address: Department of Biology, Brawijaya University, Jl. Veteran, Malang, East Java, Indonesia.

Received: 23 June 2015 Accepted: 26 August 2015
Published online: 03 September 2015

References

1. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nature Rev Genet. 2009;10:691–703.
2. Kazazian HH. Mobile elements: drivers of genome evolution. Science. 2004;303:1626–32.
3. Levin HL, Moran JV. Dynamic interactions between transposable elements and their hosts. Nat Rev Genet. 2011;12:615–27.
4. Scheifele LZ, Cost GJ, Zupancic ML, Caputo EM, Boeke JD. Retrotransposon overdose and genome integrity. Proc Natl Acad Sci USA. 2009;106:13927–32.
5. Siomi H, Siomi MC. On the road to reading the RNA-interference code. Nature. 2009;457:396–404.
6. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet. 2009;10:94–108.
7. Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol. 2009;10:126–39.
8. Castel SE, Martienssen RA. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat Rev Genet. 2013;14:100–12.
9. Kawamata T, Tomari Y. Making RISC. Trends Biochem Sci. 2010;35:368–76.
10. Kuhn C-D, Joshua-Tor L. Eukaryotic Argonautes come into focus. Trends Biochem Sci. 2013;38:263–71.
11. Eichinger L, Pachebat JA, Glöckner G, Rajandream M-A, Sucgang R, Berriman M, et al. The genome of the social amoeba *Dictyostelium discoideum*. Nature. 2005;435:43–57.
12. Glöckner G, Szafranski K, Winckler T, Dingermann T, Quail M, Cox E, et al. The complex repeats of *Dictyostelium discoideum*. Genome Res. 2001;11:585–94.
13. Martens H, Novotny J, Oberstrass J, Steck TL, Postlethwait P, Nellen W. RNAi in *Dictyostelium*: the role of RNA-directed RNA polymerases and double-stranded RNase. Mol Biol Cell. 2002;13:445–53.

14. Kuhlmann M, Borisova BE, Kaller M, Larsson P, Stach D, Na JB, et al. Silencing of retrotransposons in *Dictyostelium* by DNA methylation and RNAi. Nucleic Acids Res. 2005;33:6405–17.
15. Hinas A, Reimegard J, Wagner EG, Nellen W, Ambros VR, Söderbom F. The small RNA repertoire of *Dictyostelium discoideum* and its regulation by components of the RNAi pathway. Nucleic Acids Res. 2007;35:6714–26.
16. Wiegand S, Hammann C. The 5′ spreading of small RNAs in *Dictyostelium discoideum* depends on the RNA-dependent RNA polymerase RrpC and on the Dicer-related nuclease DrnB. PLOS ONE. 2013;8:e64804.
17. Wiegand S, Meier D, Seehafer C, Malicki M, Hofmann P, Schmith A, et al. The *Dictyostelium discoideum* RNA-dependent RNA polymerase RrpC silences the centromeric retrotransposon DIRS-1 post-transcriptionally and is required for the spreading of RNA silencing signals. Nucleic Acids Res. 2014;42:3330–45.
18. Beck P, Dingermann T, Winckler T. Transfer RNA gene-targeted retrotransposition of *Dictyostelium* TRE5-A into a chromosomal UMP synthase gene trap. J Mol Biol. 2002;318:273–85.
19. Siol O, Boutliliss M, Chung T, Glöckner G, Dingermann T, Winckler T. Role of RNA polymerase III transcription factors in the selection of integration sites by the *Dictyostelium* non-long terminal repeat retrotransposon TRE5-A. Mol Cell Biol. 2006;26:8242–51.
20. Schumann G, Zündorf I, Hofmann J, Marschalek R, Dingermann T. Internally located and oppositely oriented polymerase II promoters direct convergent transcription of a LINE-like retroelement, the *Dictyostelium* Repetitive Element, from *Dictyostelium discoideum*. Mol Cell Biol. 1994;14:3074–84.
21. Bilzer A, Dölz H, Reinhardt A, Schmith A, Siol O, Winckler T. The C-module-binding factor supports amplification of TRE5-A retrotransposons in the *Dictyostelium discoideum* genome. Eukaryot Cell. 2011;10:81–6.
22. Boesler B, Meier D, Förstner KU, Friedrich M, Hammann C, Sharma CM, et al. Argonaute proteins affect siRNA levels and accumulation of a novel extrachromosomal DNA from the *Dictyostelium* retrotransposon DIRS-1. J Biol Chem. 2014;289:35124–38.
23. Horn J, Dietz-Schmidt A, Zündorf I, Garin J, Dingermann T, Winckler T. A *Dictyostelium* protein binds to distinct oligo(dA)•oligo(dT) DNA sequences in the C-module of the retrotransposable element DRE. Eur J Biochem. 1999;265:441–8.
24. Geier A, Horn J, Dingermann T, Winckler T. A nuclear protein factor binds specifically to the 3′-regulatory module of the long-interspersed-nuclear-element-like *Dictyostelium* repetitive element. Eur J Biochem. 1996;241:70–6.
25. Winckler T, Trautwein C, Tschepke C, Neuhauser C, Zündorf I, Beck P, et al. Gene function analysis by amber stop codon suppression: CMBF is a nuclear protein that supports growth and development of *Dictyostelium* amoebae. J Mol Biol. 2001;305:703–14.
26. Dingermann T, Reindl N, Brechner T, Werner H, Nerke K. Nonsense suppression in *Dictyostelium discoideum*. Dev Genetics. 1990;11:410–7.
27. Winckler T, Iranfar N, Beck P, Jennes I, Siol O, Baik U, et al. CbfA, the C-module DNA-binding factor, plays an essential role in the initiation of *Dictyostelium discoideum* development. Eukaryot Cell. 2004;3:1349–58.
28. Schmith A, Groth M, Ratka J, Gatz S, Spaller T, Siol O, et al. Conserved gene-regulatory function of the carboxy-terminal domain of dictyostelid C-module-binding factor. Eukaryot Cell. 2013;12:460–8.
29. Windhof IM, Dubin MJ, Nellen W. Chromatin organisation of transgenes in *Dictyostelium*. Pharmazie. 2013;68:595–600.
30. Caterina MJ, Milne JLS, Devreotes PN. Mutation of the third intracellular loop of the cAMP receptor, cAR1, of *Dictyostelium* yields mutants impaired in multiple signaling pathways. J Biol Chem. 1994;269:1523–32.
31. Thomson T, Lin H. The biogenesis and function of PIWI proteins and piRNAs: Progress and prospect. Annu Rev Cell Dev Biol. 2009;25:355–76.
32. Saito K, Siomi MC. Small RNA-mediated quiescence of transposable elements in animals. Dev Cell. 2010;19:687–97.
33. Moran JV, DeBerardinis RJ, Kazazian Jr HH. Exon shuffling by L1 retrotransposon. Science. 1999;283:1530–4.
34. Stajdohar M, Jeran L, Kokosar J, Blenkus D, Janez T, Kuspa A et al. dictyExpress: visual analytics of NGS gene expression in Dictyostelium. 2015. https://www.dictyexpress.org.
35. Avesson L, Reimegård J, Wagner EG, Söderbom F. MicroRNAs in Amoebozoa: deep sequencing of the small RNA population in the social amoeba *Dictyostelium discoideum* reveals developmentally regulated microRNAs. RNA. 2012;18:1771–82.
36. Veltman DM, Akar G, Bosgraaf L, van Haastert PJM. A new set of small, extrachromosomal expression vectors for *Dictyostelium discoideum*. Plasmid. 2009;61:110–8.
37. Siol O, Spaller T, Schiefner J, Winckler T. Genetically tagged TRE5-A retrotransposons reveal high amplification rates and authentic target site preference in the *Dictyostelium discoideum* genome. Nucleic Acids Res. 2011;39:6608–19.
38. Hentschel U, Zündorf I, Dingermann T, Winckler T. On the problem of establishing the subcellular localization of *Dictyostelium* retrotransposon TRE5-A proteins by biochemical analysis of nuclear extracts. Anal Biochem. 2001;296:83–91.
39. Lucas J, Bilzer A, Moll L, Zündorf I, Dingermann T, Eichinger L, et al. The carboxy-terminal domain of *Dictyostelium* C-module-binding factor is an independent gene regulatory entity. PLoS One. 2009;4(4):e5012.

7. Diskussion

In der *Dictyostelium-discoideum*-Mutante JH.D, in welcher der C-Modul-bindende Faktor A (CbfA) nur schwach exprimiert wird, zeigte sich eine deutliche Fehlregulation von Transposons (Abbildung 7.1).

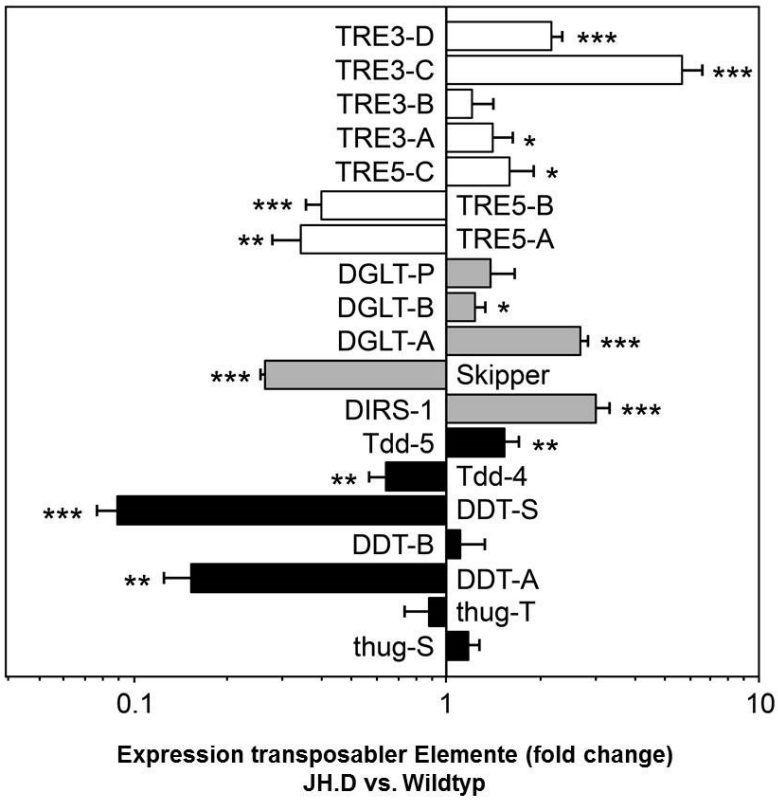


Abb. 7.1: Darstellung der Transposonfehlregulation in der Mutante JH.D im Vergleich zum Wildtyp. Es wurden die RNA-Seq-Daten aus Manuskript 2 durch Mapping der *Reads* auf die Consensus-Sequenzen der dargestellten Transposons neu ausgewertet. Abgebildet ist der *fold change* der Transposonexpressionen JH.D/Wildtyp. Eine Expression von 1 entspräche der Wildtyp-Expression; Expressionen größer 1 entsprechen einer Überexpression in der Mutante. Weiße Balken: Non-LTR-Transposons; graue Balken: LTR-Transposons; schwarze Balken: DNA-Transposons. Ergebnisse stellen Mittelwerte von drei unabhängigen Kulturen dar ± Standardabweichung. ∗: $p<0,05$; ∗∗<0,01; ∗∗∗<0,001, in Bezug zum WT (Student`s T-Test). (Modifiziert nach Manuskript 4).

Eine mögliche Abhängigkeit der Regulation mobiler Elemente durch CbfA betrifft sowohl DNA-Transposons als auch die Retrotransposons aus *D. discoideum*, darunter auch TRE5-A und DIRS-1. TRE5-A ist dabei in JH.D unterexprimiert und DIRS-1 überexprimiert (Abb. 7.1). Beide Retrotransposons sind in *D. discoideum* bereits sehr gut charakterisiert und dienen in dieser Arbeit als Modelle für die Transposonregulation in *D. discoideum.* Im Fall von TRE5-A konnte bereits gezeigt werden, dass die Regulation dieses Retrotransposons in Abhängigkeit

von CbfA nicht durch eine Bindung von CbfA an die beiden vorhandenen Promotoren (A- und C-Modul) erfolgt (Bilzer et al., 2011). Dies bedeutet, dass die *In-vitro*-Bindung von CbfA an das C-Modul und die Regulation von TRE5-A nicht in einem kausalen Zusammenhang stehen. Daher wurde ein anderer, indirekter Regulationsweg vermutet. Von TRE5-A und DIRS-1 ist bekannt, dass *Sense*- und *Antisense*-Transkripte gebildet werden (Rosen et al., 1983; Zuker et al., 1983; Schumann et al., 1994; Wiegand et al., 2013). Diese Tatsache führte zu der Annahme, dass posttranskriptionelle Stilllegungsprozesse wie RNA-Interferenz (RNAi) an der Regulation dieser Retrotransposons beteiligt sein könnten. Warum das Fehlen des Transkriptionsregulators CbfA zu einer Fehlregulation der Transposons in *D. discoideum* führt, war bislang unklar und konnte in dieser Arbeit zum Teil aufgeklärt werden.

7.1 Beeinflussung der Expression von TRE5-A durch wirtseigenen Transkriptionsregulator

Wie schon in Abbildung 7.1 dargestellt und auch in früheren Arbeiten beobachtet (Beck, 2002), fällt die Transkriptmenge von TRE5-A drastisch ab, sobald CbfA wie im Stamm JH.D unterexprimiert ist. TRE5-A gehört zu den *tRNA gene-targeted retroelements* (TREs), den Non-LTR-Retrotransposons aus *D. discoideum*, welche Integrationsorte in definierten Abständen zu tRNA-Genen vorweisen (Glöckner et al., 2001; Winckler et al., 2005). CbfA bindet an das C-Modul von TRE5-A *in vitro* (Geier et al., 1996). Da es sich bei CbfA um einen zelleigenen generellen Transkriptionsregulator handelt (Lucas et al., 2009), wurde vermutet, dass CbfA als Wirtsfaktor die Expression der Transkripte von TRE5-A unterdrückt und somit eine Retrotransposition dieses Elements verhindert. Doch die Ergebnisse dieser Arbeit zeigen, dass diese Annahme falsch war und CbfA im Gegenteil die Akkumulation der TRE5-A-Transkripte und die Retrotransposition fördert (siehe Manuskript 1). So konnte im Stamm JH.D mit stark verminderter CbfA-Expression (Winckler et al., 2001) ein drastischer Abfall von sowohl *Sense*- als auch *Antisense*-RNA von TRE5-A detektiert werden. Dieser Effekt ließ sich in Komplementationsstudien mit Expression von rekombinantem CbfA beziehungsweise rekombinanter carboxyterminaler Domäne (CTD) von CbfA in JH.D revertieren. Somit konnte gezeigt werden, dass tatsächlich CbfA beziehungsweise auch allein die CTD von CbfA für die Transkriptakkumulation von TRE5-A nötig ist. Der Abfall der Transkripte ließ auf einen direkten Einfluss von CbfA auf die beiden Promotoren von TRE5-A für die Expression der RNAs schließen. Dies konnte jedoch widerlegt werden (siehe Manuskript 1). Da sowohl *Sense*- als auch *Antisense*-RNA von TRE5-A synthetisiert werden, stellen die TRE5-A-Transkripte ein mögliches Ziel für die posttranskriptionelle Regulation durch RNA-Interferenz-Prozesse (RNAi) dar. Beeinflusst CbfA eine oder mehrere

Komponenten der RNAi-Maschinerie in *D. discoideum* und fördert dadurch die Akkumulation der TRE5-A-Transkripte? Tatsächlich ergaben vergleichende RNA-Seq-Analysen zwischen Wildtyp und JH.D (siehe Manuskript 2) eine differentielle Expression dreier Argonauten-Proteine (Agn) und einer RNA-abhängigen RNA-Polymerase (Rrp) im CbfA schwach exprimierenden Stamm JH.D.

7.2 TRE5-A- und DIRS-1-Regulation über Komponenten der RNA-Interferenz (RNAi)

TRE5-A

Im Laufe der Untersuchungen zur Funktionsaufklärung von CbfA wurden RNA-Seq-Experimente durchgeführt. Deren Auswertung ergab, dass in der CbfA-Mutante JH.D *agnC* etwa 230-fach und *agnE* etwa 3-fach im Vergleich zum Wildtyp überexprimiert wird. Die anderen Argonauten-Proteine und die RNA-abhängigen RNA-Polymerasen sind in JH.D im Vergleich zum Wildtyp nicht überexprimiert. *agnA* und *rrpC* sind dagegen in JH.D schwächer exprimiert als im Wildtyp. Diese Ergebnisse konnten mit Hilfe von qRT-PCR-Experimenten bestätigt werden (Abbildung 7.2). Zusätzlich wurden siRNAs mit einer Länge von 20 bis 22 Nukleotiden entdeckt, die TRE5-A zugeordnet werden konnten (Hinas et al., 2007). Auch dies spricht für eine TRE5-A-Regulation durch RNAi-Prozesse. Es wurde daher überprüft, ob und welche Komponenten der RNAi-Maschinerie TRE5-A regulieren.

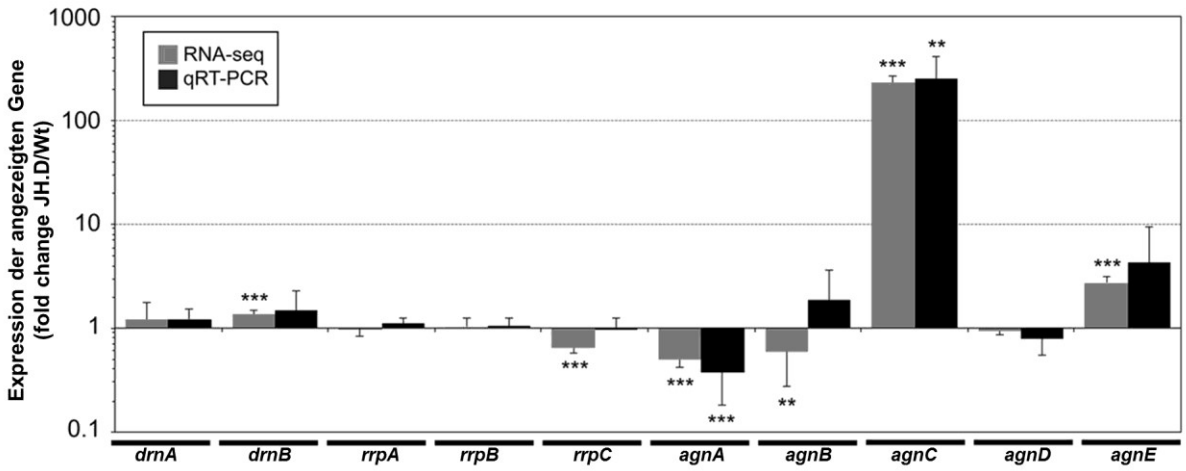


Abb. 7.2: Expression verschiedener RNAi-Komponenten im Vergleich JH.D- zu Wildtypzellen. Abgebildet ist der *fold change* der Expressionen JH.D/Wildtyp. Eine Expression von 1 entspräche der Wildtyp-Expression;

Expressionen größer 1 entsprechen einer Überexpression in der Mutante. Graue Balken: RNA-Seq dreier unabhängiger Kulturen von WT und JH.D, n = 3. Schwarze Balken: qRT-PCR der drei Kulturen der RNA-Seq und dreier zusätzlicher unabhängiger Kulturen, n = 6. Ergebnisse stellen Mittelwerte von drei bzw. sechs unabhängigen Kulturen dar ± Standardabweichung. **p<0.01, ***p<0.001, # p = 0.17, in Bezug zum WT (Student`s T-Test). (Modifiziert nach Manuskript 4).

Um testen zu können, ob Komponenten der RNAi-Maschinerie aus *D. discoideum* Einfluss auf TRE5-A haben, wurden in dieser Arbeit zunächst *Knockout*-Mutanten verschiedener RNAi-Komponenten mittels qRT-PCR auf die Expression von TRE5-A getestet. Und wie bereits vermutet, zeigten nur *agnC*$^-$- und *agnE*$^-$-Zellen eine deutliche Überexpression von TRE5-A (siehe Manuskript 4). Wurden im jeweiligen *Knockout*-Hintergrund *agnC* beziehungsweise *agnE* überexprimiert, so gleicht sich die TRE5-A-Expression in etwa dem Wildtypniveau an. *agnC* und *E* sind demnach tatsächlich an der Expressionsregulation von TRE5-A beteiligt. Die Ergebnisse wurden allerdings mit Hilfe des CbfA-Unterexpressionsstamms JH.D ermittelt. CbfA ist jedoch vermutlich auch für die Regulation der Retrotransposition beziehungsweise Integration von TRE5-A verantwortlich (siehe Abschnitt 7.4). Daher wurden *agnC* und *agnE* im Wildtyp überexprimiert. Dies hat den positiven Effekt, dass die CbfA-Expression nicht verändert ist. Gleichzeitig lässt sich überprüfen, ob die Überexpression von *agnC* und *agnE* zu einer Verringerung der TRE5-A-Expression führen.

Die Überexpression von *agnC* und *agnE* erfolgte durch Erzeugung der so genannten Genaktivierungsstämme (GA). Mit Hilfe homologer Rekombination wird der geneigene Promotor gegen einen konstitutiven Actin-15-Promotor ausgetauscht, wobei die umliegenden Genbereiche nicht verändert werden (Abbildung 7.3). Dies ist von Vorteil, da anschließend weitere Transformationen möglich sind. Und die sonst verwendeten Expressionsplasmide können zu vielen Insertionen an verschiedenen unbekannten Positionen im Genom führen, welche anormale Genexpressionen erzeugen und weiterführende Experimente verfälschen könnten (Windhof et al., 2013).

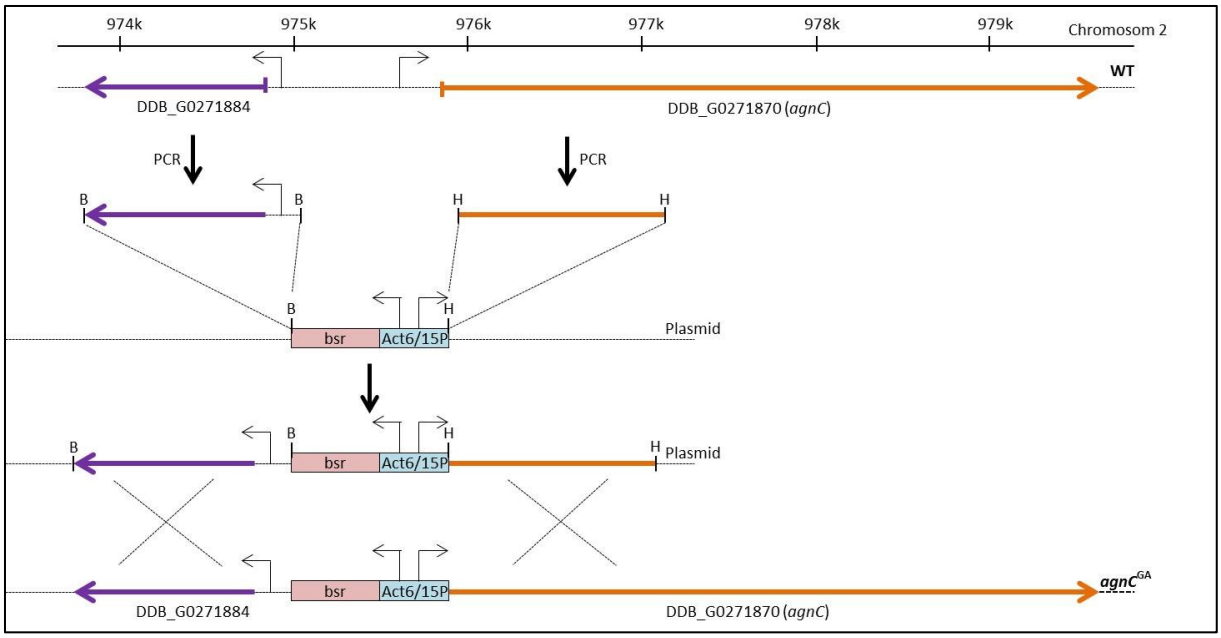


Abb. 7.3: Methode der Genaktivierung am Beispiel *agnC*. Dargestellt ist der genomische *agnC*-Lokus in *D. discoideum*. Durch homologe Rekombination von Genaktivierungsplasmid und genomischer DNA erfolgt die *agnC*-Expression durch den konstitutiv aktiven Actin15-Promotor. PCR: Polymerasekettenreaktion; B: BamHI-Schnittstelle; H: HindIII-Schnittstelle; bsr: Blasticidinresistenzkassette; Act6/15P: Actin6/Actin15-Promotor – Hybrid; Pfeile: Transkriptionsrichtung. (siehe Manuskript 4 für weitere Informationen).

Die GA-Stämme, welche entweder AgnC oder AgnE überexprimieren, zeigten eine signifikante Reduktion der TRE5-A-Transkripte. Im AgnC-GA-Stamm konnte ebenfalls eine signifikante Abnahme der Retrotranspositionsfrequenz von TRE5-A verzeichnet werden. Die Abnahme der Retrotransposition korreliert sehr gut mit der Abnahme der Transkriptmenge: AngC und AgnE werden vermutlich über die siRNAs von TRE5-A zu den Transkripten geführt und sorgen an dieser Stelle für die Degradation der Element-RNA. Dies führt wiederum dazu, dass keine TRE5-A-Proteine synthetisiert werden können, welche für die Transposition benötigt werden. Im Fall von TRE5-A konnte somit zum ersten Mal gezeigt werden, dass die Regulation dieses Retrotransposons über Komponenten des zelleigenen RNAi-Systems erfolgt.

DIRS-1

D. discoideum scheint verschiedene Strategien für die Regulation verschiedener Transposons entwickelt zu haben. So wird DIRS-1 von AgnA reguliert: In *agnA*$^-$-Stämmen nimmt die Menge an DIRS1-siRNAs ab und DIRS-1-Transkripte können akkumulieren (Boesler et al., 2014). Auch ist die RNA-abhängige RNA-Polymerase RrpC an der DIRS-1-Regulation beteiligt, wie bei der Erzeugung sekundärer siRNAs, welche das Stilllegungssignal verstärken (siehe Manuskript 3). Darauf deuten auch die Ergebnisse der

erwähnten RNA-Seq-Experimente hin: Dort zeigte sich in JH.D eine Unterexpression von *agnA* und *rrpC* und eine gleichzeitige Überxpression von DIRS-1 (siehe Abb. 7.1 und 7.2). Rrps sind an der TRE5-A-Regulation nicht beteiligt (siehe Manuskript 4). Ein weiterer Unterschied zwischen beiden Transposons zeigt sich in der Menge der siRNAs. So existiert eine sehr hohe Anzahl von DIRS-1-siRNAs, die bis zu 20 % aller detektierbaren siRNAs in *D. discoideum* ausmachen, während die Menge von TRE5-A-generierter siRNAs nur 0,05 % beträgt (Hinas et al., 2007; Manuskript 3). Ausschließen kann man bei beiden Transposons eine Beteiligung des Dicer-Homologs DrnB an der siRNA-Produktion und der generellen Transposonregulation, da der vorhandene *Knockout*-Stamm nicht zu entsprechenden Ergebnissen wie veränderter siRNA-Menge oder Expression führt. Dies scheint folgerichtig, da DrnB in der Zelle in erster Linie an der Bildung von miRNAs beteiligt zu sein scheint (Avesson et al., 2012). Über die Funktion von DrnA, ebenfalls ein Dicer-Homolog, kann nur spekuliert werden, da bisher noch keine entsprechenden *Knockout*-Stämme vorhanden sind, um die Proteinfunktionen zu studieren. Vermutlich ist dieses Protein im Fall von TRE5-A und DIRS-1 für die Bildung der primären siRNAs nötig, um die nachfolgenden RNAi-Prozesse zu aktivieren.

Warum aber betreibt die Zelle im Fall von DIRS-1 einen vermeintlich höheren Aufwand, um die Akkumulation von Transkripten und damit Transposition von DIRS-1 zu verhindern? DIRS-1 ist mit 235 vorhandenen Kopien das in *D. discoideum* am häufigsten vorkommende Transposon (Glöckner et al., 2001; Eichinger et al., 2005). Die Transposition erfolgt stets in bereits vorhandene Kopien an jeweils einem Ende der Chromosomen, wo diese DIRS-1-Cluster wahrscheinlich die Centromere ausbilden (Eichinger et al., 2005; Glöckner und Heidel, 2009). Eine unkontrollierte Vermehrung von DIRS-1 und damit ständige Integration in die bereits vorhandenen Cluster würde vermutlich zu einer Instabilität der Centromere führen. Da Centromere für die exakte Aufteilung der Chromosomen auf die Schwesterzellen nötig sind, würde deren Instabilität schließlich in einer Instabilität des gesamten Genoms resultieren. Wird durch Stilllegung von DIRS-1 die Größe der Centromere selektioniert? Würde DIRS-1, wenn es an einer beliebigen Position auf den Chromosomen integriert, zu der Ausbildung einer neuen centromeren Region führen? Und reicht dafür ein Element oder wird eine bestimmte Menge an Elementen und anderen Sequenzen benötigt, wie das beispielsweise auch in *Zea mays* der Fall ist? Dort wird eine bestimmte Anzahl an transposablen Elementen benötigt, die eine Domäne ausbilden, welche für die Funktionalität der Centromere entscheident ist (Jin et al., 2005). Dabei wird diese Domäne nicht nur von einem transposablen Element gebildet (Jin et al., 2005). Dies korreliert wiederum mit den Kenntnissen, dass in *D. discoideum* auch Skipper in centromere Bereiche integriert (Glöckner und Heidel, 2009). Die unkontrollierte Vermehrung von DIRS-1-Elementen wäre zusammenfassend vermutlich letal für die Zelle. TRE5-A dagegen integriert spezifisch in die

Nähe von tRNA-Genen (Marschalek et al., 1989; Hofmann et al., 1991; Marschalek et al., 1992), die vermutlich genomisch „sichere“ Integrationsorte darstellen.

7.3 Förderung der TRE5-A-Expression durch zelleigene Suppression von Komponenten des RNAi-Systems

Die Auswertung der RNA-Seq-Experimente ergab, dass CbfA die Argonauten-Proteine AgnC und AgnE supprimiert. Dies bedeutet, dass beide womöglich CbfA-abhängig reguliert werden. Es konnte gezeigt werden, dass beide Argonauten-Proteine die TRE5-A-Expression unterdrücken und AgnC auch die Retrotransposition von TRE5-A reduziert (siehe Manuskript 4). Ebenso fördert CbfA Expression und Retrotransposition von TRE5-A (siehe Manuskript 1). Im Laufe dieser Arbeit konnte gezeigt werden, dass beide Prozesse in direktem Zusammenhang stehen. So supprimiert CbfA die Argonauten-Proteine AgnC und AgnE und beeinflusst damit TRE5-A positiv, indem es die Funktion der RNAi-Maschinerie in der Regulation von TRE5-A unterdrückt.

An dieser Stelle muss erwähnt werden, dass die Regulation von AgnC und AgnE durch CbfA möglicherweise über verschiedene Wege erfolgt. Komplementationsstudien, in welchen rekombinantes CbfA oder rekombinante CTD von CbfA in JH.D exprimiert werden, führten nur im Fall von AgnC zu einer fast vollständigen Komplementation (siehe Manuskript 4). Im Fall von AgnE konnte keine statistisch signifikante Reduzierung der Überexpression durch CbfA ermittelt werden. Jedoch ist erwähnenswert, dass komplettes rekombinantes CbfA die *agnE*-Überexpression immerhin um 52 % senkt. Durch die CTD erfolgt keine Komplementation. An dieser Stelle kann spekuliert werden, ob die verschiedenen Proteindomänen von CbfA für diese unterschiedliche Regulation von AgnC und AgnE durch CbfA verantwortlich sind oder ob CbfA überhaupt an der AgnE-Regulation beteiligt ist.

Nach Betrachtung der Ergebnisse stellt sich die Frage, warum CbfA als ein zelleigener allgemeiner Transkriptionsregulator den AgnC/AgnE-abhängigen Weg des zelleigenen RNAi-Systems unterdrückt und somit TRE5-A-Expression und –Retrotransposition unterstützt. RNAi evolvierte womöglich als ein Abwehrprozess gegen Viren und Transposons und die ersten entdeckten siRNAs stammten von Viren (Hamilton und Baulcombe, 1999; Baulcombe, 2004). Viren haben einige Abwehrmechanismen entwickelt, die es ihnen ermöglichen, dass Wirts-RNAi-System zu umgehen beziehungsweise zu hemmen. So findet beispielsweise im Fall des *Flock house virus* aus *Drosophila melanogaster* die virale Replikation in Membranvesikeln statt und die Zwischenstufen sind dem RNAi-System somit nicht zugänglich (Miller et al., 2001; Kopek et al., 2007; Venter und Schneemann, 2008). Auch

kodieren Viren für Suppressoren des RNAi-Systems, die zum Beispiel an dsRNA binden und Dicer-Proteine inhibieren (van Rij, 2006). Für Transposons beziehungsweise Retrotransposons sind solche Strategien nicht bekannt. Jedoch zeigen einige Studien in *Drosophila* und *Arabidopsis*, dass kleine RNAs (sRNA) von transposablen Elementen (TE) in der Lage sind, die Expression von Wirtsgenen zu beeinflussen und zwar durch partielle Sequenzübereinstimmungen zwischen TE-sRNA und mRNA des entsprechenden Wirtgens (Rouget et al., 2010; McCue et al., 2013). In *Arabidopsis* konnte gezeigt werden, dass TEs damit in der Lage sind, Mechanismen zu supprimieren, die für die Stilllegung von TEs verantwortlich sind (McCue et al., 2013). Diese Abwehrmechanismen existieren allerdings unabhängig vom RNAi-System in *Arabidopsis* und können mit den in dieser Arbeit erhaltenen Daten schwer verglichen werden. In *D. discoideum* wird DIRS-1 erfolgreich durch RNAi supprimiert, die TRE5-A-Suppression durch RNAi jedoch durch CbfA unterbunden. So lässt sich spekulieren, ob TRE5-A in irgendeiner Form möglicherweise positive Effekte auf die Zelle hat. So ist in *Saccharomyces cerevisiae* kein RNAi-System detektierbar und der Verlust des RNAi-Systems scheint positiv für *S. cerevisiae* zu sein. So kann ein sogenanntes Killervirus von der Zelle beherbergt werden, welches dazu führt, dass Killer-positive Stämme das Wachstum Killer-negativer Stämme inhibieren, was zu einem klaren Vorteil führt (Welsh und Leibowitz, 1982; Billmyre et al., 2013). Durch Expression verschiedener RNAi-Komponenten wie Dicer- und Argonauten-Proteine in *S. cerevisiae* kann das RNAi-System erfolgreich aktiviert werden (Drinnenberg et al., 2009). Der Effekt des Killervirus wird tatsächlich unterbunden, wenn das RNAi-System in *S. cerevisiae* aktiviert wird (Drinnenberg et al., 2009). Daher ist der Verlust von RNAi in *S. cerevisiae* für diesen Organismus von Vorteil. Da TRE5-A sehr spezifisch in die Nähe von tRNA-Genen integriert (Marschalek et al., 1989; Marschalek et al., 1992; Hofmann et al., 1991) und die Integration dort vermutlich nicht zu Veränderungen anderer Gene führt, wird dies womöglich keine Vorteile für die Zelle bringen. Daher ist es unwahrscheinlich, dass *D. discoideum* über CbfA TRE5-A positiv reguliert, um zum Beispiel Wachstums- und Entwicklungsvorteile zu erhalten. Allerdings könnte die erfolgreiche Retrotransposition von TRE5-A *D. discoideum* ein gewisses Maß an Genomflexibilät schaffen, die positiv für die Zelle wäre. Es wäre aber auch möglich, dass die positive Regulation von TRE5-A von der Zelle nur toleriert wird und die Kontrolle einer Komponente des RNAi-Systems aus vermutlich evolutionärem Druck das eigentliche Ziel der Zelle gewesen ist. Dadurch wurde die TRE5-A-Supprimierung durch RNAi aufgehoben. Da TRE5-A jedoch in „sichere“ Orte integriert, könnte dies von der Zelle in Maßen tolerierbar sein. Wieso sollte die Zelle aber die *agnC*-Expression auf geringem Niveau halten beziehungsweise warum besitzt dieses Gen einen starken Promotor, wenn es nicht exprimiert werden soll? Die Funktion von AgnC in *D. discoideum* ist bisher völlig unbekannt. Die Expression dieses Gens ist sowohl im vegetativen Stadium als auch während der

Entwicklung sehr gering (Stajdohar et al., 2015). *cbfA* hingegen wird stärker exprimiert und das Proteinlevel bleibt auch während der Entwicklung immer konstant (Stajdohar et al., 2015; unpublizierte Daten). Ob eine der Funktionen von CbfA die Supprimierung von AgnC ist, bleibt trotzdem spekulativ. Es wäre entscheidend herauszufinden, welche Gene von AgnC noch reguliert würden, wenn es nicht durch CbfA supprimiert wird. Vergleichende RNA-Seq-Analysen zwischen Wildtyp-, JH.D- und GA-AgnC-Stämmen könnten darüber erste Informationen bieten. Ob CbfA *agnC* direkt durch Bindung an dessen Promotor reguliert, könnte wie in Manuskript 1 mit Hilfe eines Luciferase-Assays gestestet werden. Besteht eine Regulation des *agnC*-Promotors, sollte ein signifikanter Unterschied in der Promotoraktivität zwischen Wildtyp und CbfA-unterexprimierendem Stamm JH.D detektierbar sein.

7.4 Beeinflussung der Retrotransposition von TRE5-A durch CbfA

CbfA beeinflusst die Retrotransposition von TRE5-A (siehe Manuskript 1), scheinbar noch auf anderem Weg ohne Regulation des RNAi-Systems. Die Verminderung der CbfA-Expression führt in JH.D zu einer Reduktion der Retrotransposition von TRE5-A um mehr als 90 %. Revertiert werden kann dieser Effekt ausschließlich durch Expression eines kompletten CbfA-Proteins. Die CTD von CbfA allein ist nicht befähigt, die verminderte Retrotransposition aufzuheben. Im Fall der TRE5-A-Expression dagegen ist die CTD allein ausreichend, um die Menge an TRE5-A-Transkripten zu erhöhen. Werden daher Expression und Retrotransposition von TRE5-A durch unterschiedliche Funktionen von CbfA reguliert? Da das komplette CbfA für die Retrotransposition nötig ist, lässt sich über eine Rolle der JmjC/ZF-Region in diesem Prozess spekulieren. CbfA ist ein DNA-bindendes Protein. Die TRE5-A-Transposition erfolgt jedoch vermutlich über die *target-primed* r*everse transcription* (Manuskript 1). Dabei entsteht jedoch keine dsDNA-Zwischenstufe, an die CbfA binden könnte (siehe Einleitung 3.2 und 3.2.1). Für die JmjC/ZF-Region wird eine Chromatin-umgestaltende Funktion vermutet. Vielleicht führt eine durch CbfA regulierte Änderung des Chromatins an den tRNA-Genen als Intergrationsort für TRE5-A zu einer vereinfachten Integration? Die zweite Zinkfinger-Region von CbfA besitzt Ähnlichkeit zum *Plant-homeo-domain*-Typ von Proteinen, die an Nukleosomen binden (Bienz, 2005; Sanchez und Zhou, 2011). Nukleosomen sind eine Organisationsstruktur des Chromatins. Dabei sind 147 bp DNA um ein Histon-Oktamer gewickelt (Richmond und Davey, 2003). Benachbarte Nukleosomen sind über einen 10 bis 50 bp langen DNA-Abschnitt miteinander verbunden (van Holde, 1989). tRNA-Gene sind in Regionen eingebettet, die nur eine geringe Nukleosomenverbreitung aufweisen, wahrscheinlich bedingt durch ihre essentielle Rolle in der Translation oder durch die Bindung der RNA-Polymerase III (Westenberger et al., 2009).

Außerdem ist bekannt, dass die Position von Nukleosomen durch DNA-bindende oder Chromatin-modellierende Proteine verändert werden kann (Vinayachandran et al., 2009). Kann demnach TRE5-A in eine Region zwischen Nukleosom und RNA-Polymerase III integrieren, weil CbfA womöglich die Position von Nukleosomen verändert? Die Nukleosomenposition und deren mögliche Änderung lassen sich experimentell nachweisen. Platt et al. haben 2013 in *D. discoideum* eine Methode etabliert, welche die genaue Position von Nukleosomen ermitteln kann: Dabei nutzt man die Micrococcal-Nuklease, die präzise in dem DNA-Abschnitt zwischen zwei Nukleosomen schneidet. Die freigesetzte, durch Nukleosomen geschützte DNA wird dann sequenziert und die erhaltenenen Sequenzen aufs Genom kartiert. Mit Hilfe dieser Methode können Nukleosompositionen von Wildtyp und Mutanten verglichen werden. Sollte in JH.D eine zum Wildtyp veränderte Position der Nukleosomen an tRNA-Genen festgestellt werden, die auch rein durch die Integration von TRE5-A resultieren könnte, müsste geprüft werden, ob CbfA vielleicht an die Histone der Nukleosomen bindet. Dies könnte durch Hybridisierungsexperimente getestet werden. Zu allererst muss jedoch die Funktion der JmjC/ZF-Region von CbfA aufgeklärt werden, ob diese wirklich Chromatin-verändernde Aufgaben ausübt.

7.5 Funktion des Transkriptionsregulators CbfA

Entdeckt wurde der C-Modul-bindende Faktor A (CbfA) aus *D. discoideum* durch seine Bindung an das C-Modul des Retrotransposons TRE5-A *in vitro* (Geier et al., 1996). Um mehr über die Funktion von CbfA in der Zelle zu erfahren, wurde die CbfA-Supressionsmutante JH.D erstellt, welche einen starken Entwicklungsdefekt aufweist (Winckler et al., 2001). Zunächst konnte mit *Microarray*-Analysen von Wildtyp und JH.D gezeigt werden, dass es sich bei diesem Protein um einen allgemeinen Transkriptionsregulator handelt, der in der Zelle viele verschiedene Gene positiv beziehungsweise negativ beeinflusst, wobei die CTD von CbfA als eine eigenständige genregulatorische Einheit dargestellt werden konnte (Lucas et al., 2009). Im Vergleich zu *Microarray*-Analysen können die *Next-Generation-Sequencing*-Technologien qualitativ und quantitativ eine weitaus höhere Sensitivität und Genauigkeit vorweisen. Um noch mehr über die Funktion von CbfA aussagen zu können, wurde in dieser Arbeit mittels RNA-Seq noch einmal die Genexpression von Wildtyp und JH.D verglichen. Mit Hilfe dieser Methode konnten 1030 CbfA-abhängige Gene identifiziert werden, welche mindestens eine dreifache Änderung der Genexpression im Vergleich zum Wildtyp aufweisen. Von diesen 1030 ermittelten Genen, waren nur 35 % auf dem *Microarray* vertreten. Dies zeigt die Notwendigkeit, die Analyse der Genexpression zur Ermittlung CbfA-abhängiger Gene mit

einer sensitiveren Methode zu wiederholen. Außerdem konnten die durch RNA-Seq erhaltenen Ergebnisse ausgewählter Gene mit Hilfe von quantitativer RT-PCR (qRT-PCR) bestätigt werden (siehe Manuskript 2). In diesen wurde die Expression verschiedener CbfA-abhängiger Gene in Zellkulturen gemessen, die unabhängig von denen sind, die für die RNA-Seq-Experimente genutzt wurden. Dies deutet auf eine gute Reproduzierbarkeit der erhaltenen Ergebnisse hin.

Die von Winckler und Kollegen 2001 ermittelten Phänotypen von JH.D wie verringerte Phagozytoserate und verlangsamtes Wachstum auf Bakterien lassen sich gut mit den ermittelten CbfA-abhängigen Genen in Beziehung bringen. So finden sich viele von CbfA positiv beeinflusste Gene, deren Proteine an Prozessen wie der Zusammensetzung von Phagozytosevesikeln, der Zusammensetzung des Zytoskeletts und allgemeinen zellulären Prozessen wie der Zellteilung beteiligt sind. Im Stamm JH.D, in welchem CbfA nur zu ca. 5 % exprimiert wird (Winckler et al., 2001), sind diese Gene unterexprimiert. Dies führt in der Folge zu den ermittelbaren Phänotypen. Allerdings liefern Gene von an Phagozytose beteiligten Proteinen in sowohl RNA-Seq, qRT-PCR als auch *Microarray*-Analysen verschiedene Ergebnisse. Da in jedem dieser Experimente verschiedene Kulturen verwendet wurden, scheint die Ursache in einer kulturabhängigen und sehr variablen Expression der betroffenen Gene zu liegen. Dies scheint plausibel, da auch der Prozess der Phagozytose in *D. discoideum* an sich ebenfalls zu einer schnellen Anpassung der Zellen und damit schnellen Änderung der Expression beteiligter Gene führt (Vogel et al., 1980; Bozzaro et al., 1987; Bracco et al., 1997; Balbo und Bazzaro, 2006; Sillo et al., 2008).

Auch wurde mittels RNA-Seq überprüft, welchen Einfluss eine konstitutive Expression der CTD von CbfA in JH.D auf die Genregulation hat. Es konnte gezeigt werden, dass ein erstaunlich hoher Anteil von CbfA beeinflussten Genen allein durch die CTD reguliert wird. So reagierten 85 % der positiv und 76,7 % der negativ von CbfA regulierten Gene auf die konstitutive Expression der CTD im JH.D-Hintergrund. Viele Gene sind demnach allein durch die CTD regulierbar, ohne dass der Rest des CbfA-Proteins dafür benötigt wird. Rund 70 % dieser CTD-regulierten Gene zeigten ein stärkeres Expressionslevel als im Wildtyp. Die Expression der entsprechenden Gene erfolgte demzufolge nicht nur wieder auf Wildtypniveau, wurde also komplementiert, sondern darüber hinaus. Da die konstitutiv exprimierte CTD in JH.D stark überexprimiert ist, was durch Western Blot bestätigt werden konnte (siehe Manuskript 4), handelt es sich dabei vermutlich um Überexpressionsartefakte.

7.5.1 Die Familie der C-Modul-bindender-Faktor-ähnlichen Proteine

In dieser Arbeit konnte eine neue Proteinfamilie definiert werden, die „C-Modul-bindender-Faktor-ähnlichen Proteine“. Sie ist definiert über eine konservierte JumonjiC/Zinkfinger 1/Zinkfinger 2 (JmjC/ZF1/ZF2)-Architektur mit variabler CTD. Dazu gehören die CbfA-Proteine der *Dictyosteliidae*, die CbfA-ähnlichen (CbfB-) Proteine der *Dictyosteliidae* und CbfA-ähnliche Proteine in Pilzen (Abbildung 7.4).

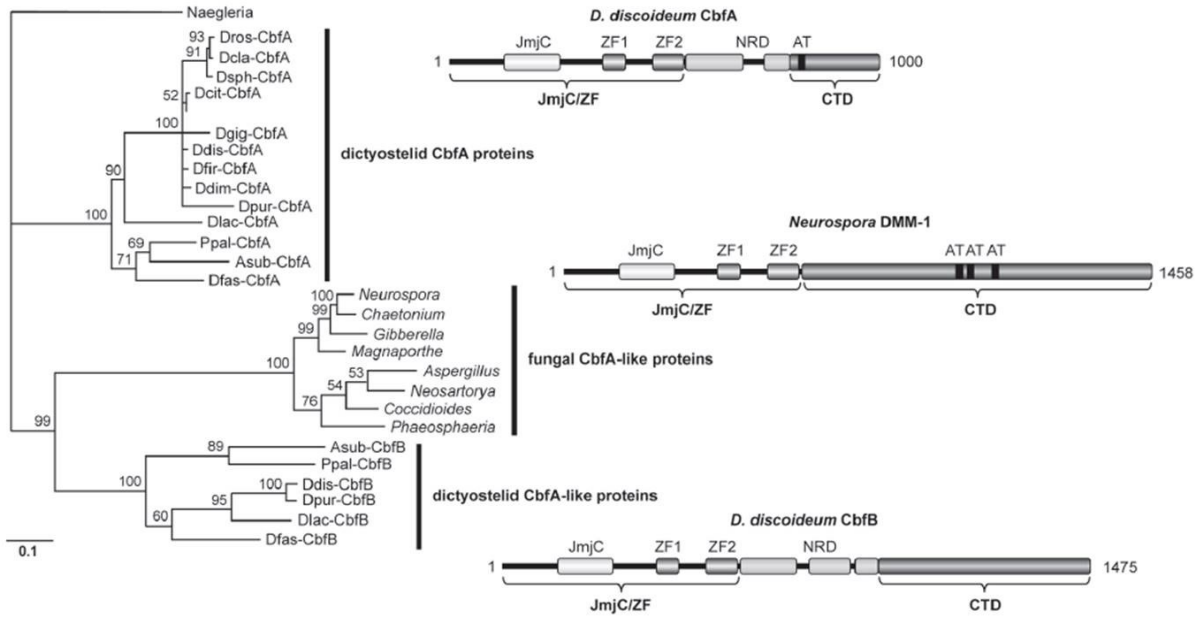


Abb. 7.4: Phylogenie der Cbf-ähnlichen Proteine und schematische Darstellung von CbfA und CbfB aus *D. discoideum* und DMM-1 aus *Neurospora*. Proteinsequenzen, die der JmjC/ZF-Region des *D. discoideum* CbfAs entsprechen wurden mit ClustalX 2.0 *aligned* und mit Tree-Puzzle 5.2 analysiert. Jmj: Jumonji-Domäne; ZF: Zinkfinger; NRD: Asparaginreiche Domäne; AT:AT-Haken; CTD: Carboxy-terminale Domäne. (siehe auch Manuskript 2 für weitere Informationen).

Die CbfA-ähnliche ZF-Region findet sich auch in Proteinen von Pflanzen und Tieren. Diese Proteine weisen dann aber nur ZF1 oder ZF2 auf. Ebenfalls konnte die typische Kombination der ZF mit einer JmjC-Domäne in Pflanzen und Tieren nicht gefunden werden. Jedoch findet sich diese Kombination in verschiedenen filamentösen Pilzen, in *Naegleria gruberi* und in *Acanthamoeba castellanii* und natürlich in allen vier Gruppen der *Dictyosteliidae* (Abb. 7.4). Phylogenetische Analysen dieser Region zeigten eine deutliche Trennung zwischen den CbfA-Proteinen der *Dictyosteliidae* und den CbfA-ähnlichen Proteinen (CbfB) der *Dictyosteliidae* und deren Orthologen in Pilzen. Die ZF2-Region ist in allen Organismen hoch konserviert. Dagegen zeigen sich in der ZF1-Region kleine Sequenzunterschiede. Diese führen zu der Annahme, dass es sich bei CbfB um die ursprüngliche Form des ZF1 handeln

könnte. So teilen sich ZF1-Regionen von CbfB-Proteinen aus *D. fasciculatum*, eine der am frühesten divergierte Art der *Dictyosteliidae*, und Pilzen ein Sequenzmotiv (siehe Manuskript 2). Eine kurze Verlängerung dieser Sequenz führte vermutlich zu einem Motiv, welches sich nur in CbfA-Proteinen der *Dictyosteliidae* finden lässt. So lässt sich zusammenfassen, dass sich die CbfA-ähnlichen Proteine vermutlich vor der Abspaltung der *Dictyosteliidae* entwickelt haben. In der Evolution der *Dictyosteliidae* entwickelten sich dann wahrscheinlich aus den CbfB-Proteinen die CbfA-Proteine.

Über die Funktionen dieser neu definierten Proteinfamilie der „C-Modul-bindender-Faktor-ähnliche Proteine“ lässt sich nur spekulieren. So ist DMM-1 (DNA Methylation Modulator-1) aus *Neurospora crassa* (Abbildung 7.4) an der Inhibierung einer Ausbreitung der Methylierung von Histon-3-Lysin-9 beteiligt (Honda et al., 2010). Dabei sind die JmjC- und ZF-Regionen für die Funktion des Proteins entscheidend und Deletionen und Mutationen in diesen Domänen führen zum Funktionsverlust (Honda et al., 2010). Die JmjC/ZF-Region von DMM-1 zeigt eine hohe Ähnlichkeit zu CbfA aus *D. discoideum*, die CTDs der beiden Proteine sind jedoch verschieden. Über eine Funktion der CTD von DMM-1 ist bislang nichts bekannt, nur dass sie für die Funktionsweise des Proteins nicht entscheidend ist (Honda et al., 2010). Ob die JmjC/ZF-Region von CbfA aus *D. discoideum* ebenfalls eine demethylierende Funktion ausübt, muss erst untersucht werden. Die hohe Sequenzähnlichkeit zu DMM-1 lässt aber darauf schließen. Ebenso dürften die anderen Proteine dieser Familie ähnliche Funktionen ausüben, da gerade die JmjC/ZF-Region eine hohe Konservierung aufweist.

7.5.2. CbfA-Proteine der *Dictyosteliidae*

Durch Strukturanalysen von CbfA mit CbfA-Proteinen anderer *Dictyosteliidae*-Arten konnte herausgefunden werden, dass die jeweiligen CTDs dieser Proteine innerhalb der *Dictyosteliidae* gut konserviert sind. Die *Dictyosteliidae* sind in den Gruppen 1 bis 4, aufgrund der Analyse der Sequenzen der kleinen Untereinheit der rRNA und dem Vergleich von α-Tubulin, eingruppiert (Schaap et al., 2006). *D. discoideum* wurde der Gruppe 4 zugeordnet. Die CTDs der CbfA-Proteine innerhalb der Gruppe 4 weisen eine Identität von bis zu 98 % auf. Verglichen mit den Gruppen 1, 2 und 3 weisen die CTDs der entsprechenden Arten noch Identitäten von mehr als 40 % auf. Des Weiteren sind die CbfA-CTDs dieser Gruppen verlängert, ohne dass diese Verlängerungen Homologien zueinander erkennen lassen. Um zu testen, ob diese Domänen auch funktionell ähnlich sind, wurde die CTD von *Polysphondylium pallidum* konstitutiv in JH.D (*D. discoideum*) exprimiert und die Änderungen

der Genexpression durch RNA-Seq ermittelt und mit der Expression der *D. discoideum*-CTD in JH.D verglichen. *P. pallidum* gehört der Gruppe 2 an und divergierte damit schon frühzeitig von den Gruppe-4-Arten wie *D. discoideum*. Erstaunlicherweise wurden dennoch 73,3 % der CbfA-aktivierten und 51,4 % der von CbfA negativ regulierten Gene von der artfremden CTD wiederhergestellt. Dies weist nicht nur auf eine strukturelle Konservierung, sondern viel mehr auch auf eine funktionelle Konservierung der CTDs der CbfA-Proteine innerhalb der *Dictyosteliidae* hin. Dies ist besonders bemerkenswert im Hinblick auf die 600 Millionen Jahre Evolutionszeit der *Dictyosteliidae* (Glöckner und Noegel, 2013). Erwähnenswert ist außerdem, dass *D. discoideum* einen höheren A/T-Gehalt im Genom aufweist als *P. pallidum* (Heidel et al., 2011). Die CTD von *D. discoideum* benötigt die Bindung an A/T-reiche Sequenzen *in vitro* und *in vivo*, um seine genregulatorische Funktion auszuführen (Horn et al., 1999; Lucas et al., 2006). Umso bemerkenswerter ist die Tatsache, dass die *P. pallidum*-CTD die Expression einer derart hohen Anzahl von Genen im Genom von *D. discoideum* wiederherstellen kann.

Innerhalb der *Dictyosteliidae* sind die CTDs der CbfA-Proteine konserviert. Jedoch sind die CTDs der CbfB-Proteine (CbfA-ähnliche) der *Dictyosteliidae* verlängert, und sie zeigen keinerlei Ähnlichkeit zueinander Dies lässt in Zusammenhang mit der Konservierung der JmjC/ZF-Region vermuten, dass sich diese CTDs evolutionär schneller entwickelt haben, vermutlich unter Art-spezifischem Anpassungsdruck.

Die Unterschiede zwischen den CbfA- und CbfB-Proteinen der *Dictyosteliidae* führen vermutlich zu Spezifikationen in ihren Proteinfunktionen. So kann im *D. discoideum*-Stamm JH.D CbfB nicht die ausgefallenen Funktionen von CbfA übernehmen, sonst könnte sich ein so ausgeprägter Phänotyp wie in JH.D vermutlich nicht ausbilden. Welche Unterschiede in den Funktionen beider Proteine aber genau liegen, muss noch weiter untersucht werden. Bisher war die Stilllegung des *cbfB*-Gens in JH.D noch nicht möglich, was jedoch Voraussetzung für die entsprechenden Experimente ist.

7.6 Ausblick

Die posttranskriptionelle Regulation von TRE5-A konnte in dieser Arbeit soweit aufgeklärt werden, dass die Argonauten-Proteine AgnC und AgnE als Komponenten des RNAi-Systems in *D. discoideum* TRE5-A regulieren. Dabei bleibt weiter zu klären, ob das Dicer-Protein DrnA für die Synthese der siRNAs von TRE5-A verantwortlich ist. Die Herstellung eines entsprechenden *drnA*$^-$-Stammes wäre für die Beantwortung dieser Fragestellung hilfreich. Auch Nachweise der Bindung von siRNA von TRE5-A in RISC-Komplexe mit

entweder AgnC oder AgnE als Kernkomponente und der Beweis des Abbaus von TRE5-A-RNA durch RISC wären für die noch detailliertere Aufklärung der Abfolge der posttranskriptionellen Regulation von TRE5-A interessant.

Die Genaktivierungsmutanten von AgnC und AgnE könnten weiterführend zum Beispiel durch RNA-Seq-Experimente charakterisiert werden, um noch mehr über die Funktion dieser Proteine zu erfahren. Vielleicht lässt sich dann besser aufklären, warum die Zelle einen Transkriptionsregulator zur möglichen Kontrolle von Komponenten des RNAi-Systems einsetzt.

DIRS-1 und TRE5-A werden durch verschiedene RNAi-Komponenten in unterschiedlicher Art und Weise reguliert. Daher scheint es interessant, auch die Regulation anderer Transposons in *D. discoideum* aufzuklären. Die Aufklärung der Funktionen der übrigen Argonauten-Proteine und deren eventuelles Zusammenspiel mit den RNA-abhängigen RNA-Polymerasen und Dicer-Proteinen könnte ebenfalls hilfreich für die Klärung der *Dictyostelium*-eigenen Transposonregulation sein. Auch bei DIRS-1 ist noch unklar, ob DrnA die siRNAs synthetisiert, die dann zur Hemmung des Transposons führen.

Diese Arbeit konnte Aufschluss über die Funktionen von CbfA und dessen CTD geben. Allerdings bleibt offen, welchen Einfluss und welche Funktionen die JmjC/ZF-Region hat. So wie die CTD in JH.D, einem Stamm mit geringer CbfA-Expression, exprimiert wurde, um die Funktion zu studieren, könnten ähnliche Experimente mit der JmjC/ZF-Region durchgeführt werden. Als möglicher Chromatin-Modellierer hätte CbfA vielleicht Einfluss auf die Nukleosomenstruktur in der Nähe von tRNA-Genen und könnte dadurch die Transposition von TRE5-A positiv beeinflussen. Daher wäre eine vergleichende Überprüfung der Nukleosomenpositionen in JH.D und Wildtyp mit Hilfe der Micrococcal-Nuklease und anschließender Sequenzierung der durch Nukleosomen geschützten DNA wahrscheinlich lohnend.

Sehr interessant scheint auch die Aufklärung der Funktion des CbfA-ähnlichen Proteins CbfB in *D. discoideum.* Sind die Aufgaben dieses Proteins ähnlich zu denen von CbfA? Besitzt auch die CTD von CbfB eigenständige Funktionen und wenn ja welche? Eine Mutante ähnlich zu JH.D, die nur geringe Mengen von CbfB exprimiert, oder ein *cbfB*$^{-}$-Stamm könnte darüber Aufschluss geben.

8. Literaturverzeichnis

Agger K., Christensen J., Cloos P.A., Helin K. (2008) The emerging functions of histone demethylases. *Curr. Opin. Genet. Dev*. **18**: 159-168.

Aspegren A., Hinas A., Larsson P., Larsson A., Soderbom F. (2004) Novel non-coding RNAs in *Dictyostelium discoideum* and their expression during development. *Nucleic Acids Res.* **32:** 4646-56.

Avesson L., Reimegard J., Wagner E.G., Söderbom F. (2012) MicroRNAs in Amoebozoa: Deep sequencing of the small RNA population in the social amoeba *Dictyostelium discoideum* reveals developmentally regulated micro-RNAs. *RNA*. **18(10):** 1771-1782.

Ayarpadikannan S., Kim H.S. (2014) The Impact of Transposable Elements in Genome Evolution and Genetic Instability and Their Implications in Various Diseases. *Genomics Inform.* **12(3):** 98-104.

Aziz R.K., Breitbart M., Edwards R.A. (2010) Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res*. **38(13):** 4207-17.

Balbo A., Bozzaro S. (2006) Cloning of *Dictyostelium* eIF6 (p27BBP) and mapping its nucle(ol)ar localization subdomains. *Eur J Cell Biol*. **85:** 1069–1078.

Bannert N., Kurth R. (2006) The evolutionary dynamics of human endogenous retroviral families. *Annu Rev Genomics Hum Genet*. **7:** 149–173.

Baulcombe D. (2004) RNA silencing in plants. *Nature*. **431:** 356–363.

Beck P. (2002) Etablierung eines *in vivo* Testsystems zur Untersuchung der Mobilität tRNA-Gen gerichteter Retrotransposons in *Dictyostelium discoideum. Dissertationsschrift.*

Belancio V.P., Deininger P.L., Roy-Engel A.M. (2009) LINE dancing in the human genome: transposable elements and disease. *Genome Med*. **1:** 97.

Bernstein E., Caudy A.A., Hammond S.M., Hannon G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*. **409:** 363–366.

Bienz M. (2005) The PHD finger, a nuclear protein-interaction domain. *Trends Biochem. Sci.* **31(1)**: 35-40.

Billmyre R.B., Calo S., Feretzaki M., Wang X., Heitman J. (2013) RNAi function, diversity, and loss in the fungal kingdom. *Chromosome Res*. **21(6-7):** 561-572.

Bilzer A., Dölz H., Reinhardt A., Schmith A., Siol O., Winckler T. (2011) The C-module-binding factor supports amplification of TRE5-A retrotransposons in the *Dictyostelium discoideum* genome. *Eukaryot. Cell.* **10(1):** 81-6.

Boeke J., Chapman K. (1991) Retrotransposition mechanisms. *Curr. Opin. Cell Biol*. **3:** 502–507.

Boeke J., Corces V. (1989) Transcription and reverse transcription of retrotransposons. *Ann. Rev. Microbiol*. **43:** 403–434.

Boesler B., Meier D., Förstner K.U., Friedrich M., Hammann C., Sharma C.M., Nellen W. (2014) Argonaute proteins affect siRNA levels and accumulation of a novel extrachromosomal DNA from the *Dictyostelium* retrotransposon DIRS-1. *J Biol Chem*. **289(51):** 35124-35138.

Bonner J.T., Lamont D. S. (2005) Behavior of cellular slime molds in the soil. *Mycologia.* **97:** 178-184.

Bozzaro S., Merkl R., Gerisch G. (1987) Cell adhesion: its quantification, assay of the molecules involved, and selection of defective mutants in *Dictyostelium* and *Polysphondylium*. *Methods Cell Biol*. **28:** 359–385.

Bracco E., Peracino B., Noegel A.A., Bozzaro S. (1997) Cloning and transcriptional regulation of the gene encoding the vacuolar/H+ ATPase B subunit of *Dictyostelium discoideum*. *FEBS Lett*. **419:** 37–40.

Cappello J., Cohen S.M., Lodish H.F. (1984) *Dictyostelium* transposable element DIRS-1 preferentially inserts into DIRS-1 sequences. *Mol Cell Biol*. **4:** 2207–2213.

Cappello J., Handelsman K., Lodish H.F. (1985) Sequence of *Dictyostelium* DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. *Cell*. **43:** 105–115.

Capy P. (2005) Classification and nomenclature of retrotransposable elements. *Cytogenet Genome Res*. **110(1-4)**: 457-461.

Capy P., Vitalis R., Langin T., Higuet D., Bazin C. (1996) Relationships between transposable elements based upon the integrase-transposase domains: is there a common acnestor? *J. Mol. Evol*. **42:** 359–368.

Cerutti H., Casas-Mollano J.A. (2006) On the origin and functions of RNA mediated silencing: from protists to man. *Curr Genet*. **50:** 81–99.

Chapman K.B., Byström A.S., and Boeke J.D. *(*1992) *Initiator methionine tRNA is essential for Ty1 transposition in yeast. Proc. Natl. Acad. Sci.* **89***: 3236–3240.*

Chung T., Siol O., Dingermann T., *Winckler T.* (2007) Protein interactions involved in tRNA gene-specific integration of *Dictyostelium discoideum* non-long terminal repeat retrotransposon TRE5-A. *Mol. Cell. Biol.* **27(24):** 8492-8501.

Cohen M.S., Cappello J., Lodish H.F. (1984) Transcription of *Dictyostelium discoideum* transposable element DIRS-1. *Mol Cell Biol.* **4(11):** 2332-2340.

Cordaux R., Batzer M.A. (2009) The impact of retrotransposons on human genome evolution. *Nat Rev Genet*. **10:** 691-703.

Cost G.J., Feng Q., Jacquier A., Boeke J.D. (2002) Human L1 element target-primed reverse transcription in vitro. *EMBO J*. **21:** 5899–5910.

Craig N.L., Craigie R., Gellert M., Lambowitz A.M. (2002) Mobile DNA II. *ASM Press Washington DC*.

Daboussi M.J., Capy P. (2003) Transposable elements in filamentous fungi. *Annu Rev Microbiol*. **57:** 275-299.

Dingermann T., Reindl N., Brechner T., Werner H., Nerke K. (1990) Nonsense suppression in *Dictyostelium discoideum*. *Dev. Genet.* **11:** 410-417.

Dogini D.B., Pascoal V.D., Avansini S.H., Vieira A.S., Pereira T.C., Lopes-Cendes I. (2014) The new world of RNAs. *Genet Mol Biol*. **37:** 285-293.

Dombroski B.A., Mathias S.L., Nanthakumar E., Scott A.F., Kazazian H.H. Jr. (1991) Isolation of an active human transposable element. *Science*. **254:** 1805–1808.

Doolittle R.F., Feng D.F. (1992) Tracing the origin of retroviruses. *Curr Top Microbiol Immunol*. **176:** 195-211.

Doolittle R.F., Feng D.F., Johnson M.S., McClure M.A. (1989) Origins and evolutionary relationships of retroviruses. *Quart. Rev. Biol*. **64:** 1–30.

Drinnenberg I.A., Weinberg D.E., Xie K.T., Mower J.P., Wolfe K.H., Fink G.R., Bartel D.P. (2009) RNAi in budding yeast. *Science*. **326:** 544–550.

Eichinger L., Pachebat J.A., Glöckner G., Rajandream M.A., Sucgang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q., Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F., Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P., Pilcher K., Chen G., Saunders D., Sodergren E., Davis P., Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N., Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C., Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I., Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R., Hauser H., James K., Quiles M., Madan Babu M., Saito T., Buchrieser C., Wardroper A., Felder M., Thangavelu M., Johnson D., Knights A., Loulseged H., Mungall K., Oliver K., Price C., Quail M.A., Urushihara H., Hernandez J., Rabbinowitsch E., Steffen D., Sanders M., Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A., Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y., Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C., Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R., Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A. (2005) The genome of the social amoeba *Dictyostelium discoideum*. *Nature.* **435:** 43-57.

Esnault C., Maestre J., Heidmann T. (2000) Human LINE retrotransposons generate processed pseudogenes. *Nat Genet*. **24:** 363–367.

Faix J., Kreppel L., Shaulsky G., Schleicher M. und Kimmel A. R. (2004) A rapid and efficient method to generate multiple gene disruptions in *Dictyostelium discoideum* using a single selectable marker and the Cre-loxP system. *Nucleic Acids Res.* **32:** e143.

Feng Q., Moran J.V., Kazazian H.H. Jr, Boeke J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*. **87:** 905–916.

Filée J., Siguier P., Chandler M. (2007) Insertion sequence diversity in archaea. *Microbiol Mol Biol Rev*. **71(1):** 121-57.

Finnegan D.J. (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet*. **5(4):** 103-107.

Fire A., Xu S., Montgomery M.K., Kostas S.A., Driver S.E., Mello C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*. **391(6669):** 806-811.

Gaudet P., Williams J., Fey P., Chisholm R. (2008) An anatomy ontology to represent biological knowledge in *Dictyostelium discoideum*. *BMC Genomics*. **9:** 130.

Geier A., Horn J., Dingermann T., Winckler T. (1996) Nuclear protein factor binds specifically to the 3´-regulatory module of the long-interspersed-nuclear-element-like *Dictyostelium* repetitive element. *Eur. J.Biochem.* **241:** 70-76.

Ghildiyal M., Zamore P.D. (2009) Small silencing RNAs: an expanding universe. *Nat Rev Genet*. **10(2):** 94–108.

Girard A., Hannon G.J. (2008) Conserved themes in small-RNA-mediated transposon control. *Trends Cell Biol*. **18(3):** 136–148.

Glöckner G., Eichinger L., Szafranski K., Pachebat J.A., Bankier A.T., Dear P.H., Lehmann R., Baumgart C., Parra G., Abril J.F., Guigo R., Kumpf K., Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A. (2002) Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. *Nature.* **418:** 79-85.

Glöckner G., Heidel A.J. (2009) Centromere sequence and dynamics in *Dictyostelium discoideum*. *Nucleic Acids Res*. **6:** 1809–1816.

Glöckner G., Szafranski K., Winckler T., Dingermann T., Quail M.A., Cox E., Eichinger L., Noegel A.A., Rosenthal A. (2001) The complex repeats of *Dictyostelium discoideum*. *Genome Res*. **11(4):** 585-94.

Goodwin T.J., Poulter R.T. (2001) The DIRS1 Group of Retrotransposons. *Mol. Biol. Evol.* **18(11):** 2067–2082.

Hamilton A.J., Baulcombe D.C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*. **286:** 950–952.

Hammond S.M., Bernstein E., Beach D., Hannon G.J. (2000) An RNA-directed nuclease mediates posttranscriptional gene silencing in *Drosophila* cells. *Nature*. **404:** 293–296.

Heidel A.J., Lawal H.M., Felder M., Schilde C., Helps N.R., Tunggal B., Rivero F., John U., Schleicher M., Eichinger L., Platzer M., Noegel A.A., Schaap P., Glöckner G. (2011) Phylogeny-wide analysis of social amoeba genomes highlights ancient originis for complex intercellular communication. *Genome Res*. **21**: 1882-1891.

Hickey D.A. (1992) Evolutionary dynamics of transposable elements in prokaryotes and eukaryotes. *Genetica*. **86:** 269–274.

Hinas A., Larsson P., Avesson L., Kirsebom A., Virtanen A., Soderbom F. (2006) Identification of the major spliceosomal RNAs in *Dictyostelium discoideum* reveals developmentally regulated U2 variants and polyadenylated snRNAs. *Eukaryot Cell.* **5:** 924-34.

Hinas A., Reimegård J., Wagner E.G., Nellen W., Ambros V.R., Söderbom F. (2007) The small RNA repertoire of *Dictyostelium discoideum* and its regulation by components of the RNAi pathway. *Nucleic Acids Res*. **35(20):** 6714-6726.

Hofmann J., Schumann G., Borschet G., Gösseringer R., Bach M., Bertling W.M., Marschalek R., Dingermann T. (1991) Transfer RNA genes from *Dictyostelium discoideum* are frequently associated with repetitive elements and contain consensus boxes in their 5' and 3'-flanking regions. *J. Mol. Biol*. **222(3):** 537-552.

Hohjoh H., Singer MF. (1996) Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J*. **15:** 630–639.

Honda S., Lewis Z.A., Huarte M., Cho L.Y., David L.L., Shi Y., Selker E.U. (2010) The DMM complex prevents spreading of DNA methylation from transposons to nearby genes in *Neurospora crassa*. *Genes Dev.* **24:** 443-454.

Horn J., Dietz-Schmidt A., Zündorf I., Garin J., Dingermann T., Winckler T. (1999) A *Dictyostelium* protein binds to distinct oligo(dA)-oligo(dT) DNA sequences in the C-module of the retrotransposable element DRE. *Eur. J. Biochem.* **265:** 441-448.

Hua-Van A., Le Rouzic A., Maisonhaute C., Capy P. (2005) Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. *Cytogenet Genome Res*. **110(1-4):** 426-440.

Jacquet M., Guilbaud R., Garreau H. (1988) Sequence analysis of the DdPYR5-6 gene coding for UMP synthase in Dictyostelium discoideum and comparison with orotate phosphoribosyl transferases and OMP decarboxylases. *Mol Gen Genet*. **211:** 441-445.

Jin W., Lamb J.C., Vega J.M., Dawe R.K., Birchler J.A., Jiang J. (2005) Molecular and functional dissection of the maize B chromosome centromere. *Plant Cell*. **17(5):** 1412-1423.

Johnson R.C., Reznikoff W.S. (1983) DNA sequences at the ends of transposon Tn5 required for transposition. *Nature*. **304:** 280-282.

Kazazian H.H. Jr., Moran J.V. (1998) The impact of L1 retrotransposons on the human genome. *Nat Genet*. **19:** 19–24.

Kazazian H.H. Jr. (2004) Mobile elements: drivers of genome evolution. *Science*. **303:** 1626-1632.

Kidwell M. (1992) Horizontal transfer. *Curr Opin Genet Dev.* **2:** 868–873.

Kim B.D. (2014) Foldback intercoil DNA and the mechanism of DNA transposition. *Genomics Inform*. **12(3):** 80-86.

Kim J.M., Vanguri S., Boeke J.D., Gabriel A., Voytas D.F. (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res*. **8(5)**: 464-478.

Kleckner N. (1990) Regulation of transposition in bacteria. *Annu Rev Cell Biol*. **6:** 297–327.

Klose R.J., Kallin E.M., Zhang Y. (2006) JmjC-domain-containing proteins and histone demethylation. *Nat. Rev. Genet*. **7**: 715-727.

Kopek B.G., Perkins G., Miller D.J., Ellisman M.H., Ahlquist P. (2007) Three-dimensional analysis of a viral RNA replication complex reveals a virus-induced mini-organelle. *PLoS Biol*. **5:** e220.

Kreppel L., Fey P., Gaudet P., Just E., Kibbe W.A., Chisholm R.L., Kimmel A.R. (2004) dictyBase: a new *Dictyostelium discoideum* genome database. *Nucleic Acids Res*. **32 (Database issue):** D332-333.

Kuhlmann M., Borisova B.E., Kaller M., Larsson P., Stach D., Na J., Eichinger L., Lyko F., Ambros V., Söderbom F., Hammann C., Nellen W. (2005) Silencing of retrotransposons in *Dictyostelium* by DNA methylation and RNAi. *Nucleic Acids Res*. **33(19):** 6405-6417.

Kulpa D.A., Moran J.V. (2005) Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Hum Mol Genet*. **14(21):** 3237-3248.

Kulpa D.A., Moran J.V. (2006) *Cis*-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat Struct Mol Biol*. **13:** 655–660.

Levin H.L., Moran J.V. (2011) Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet*. **12(9):** 615–627.

Loomis W.F., Welker D., Hughes J., Maghakian D., Kuspa A. (1995) Integrated maps of the chromosomes in *Dictyostelium discoideum. Genetics*. **141:** 147–157.

Luan D.D., Korman M.H., Jakubczak J.L., Eickbush T.H. (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell.* **72:** 595–605.

Lucas J., Bilzer A., Moll L., Zündorf I., Dingermann T., Eichinger L., Siol O., Winckler T. (2009) The carboxyterminal domain of *Dictyostelium* C-module-binding factor is an independent gene regulatory entity. *PLOS ONE*. **4(4):** e5012.

Mahillon J., Chandler M. (1998) Insertion sequences. *Microbiol Mol Biol Rev*. **62(3):** 725-774.

Maida Y., Masutomi K. (2011) RNA-dependent RNA polymerases in RNA silencing. *Biol Chem*. **392:** 299–304.

Malik H., Henikoff S., Eickbush T.H. (2000) Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res*. **10:** 1307–1318.

Malik H.S., Burke W.D., Eickbush T.H. (1999) The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol*. **16:** 793–805.

Malone C.D., Hannon G.J. (2009) Small RNAs as Guardians of the Genome. *Cell*. **136(4):** 656–668.

Marschalek R., Brechner T., Amon-Böhm E., Dingermann T. (1989) Transfer RNA genes: Landmarks for integration of mobile genetic elements in *Dictyostelium discoideum*. *Science.* **244:** 1493–1496.

Marschalek R., Hofmann J., Schumann G., Dingermann T. (1992) Two distinct subforms of the retrotransposable DRE element in NC4 strains of *Dictyostelium discoideum*. *Nucleic Acids Res.* **20(23):** 6247-6252.

Martens H., Novotny J., Oberstrass J., Steck T.L., Postlethwait P., Nellen W. (2002) RNAi in *Dictyostelium*: the role of RNA-directed RNA polymerases and double-stranded RNase. *Mol. Biol. Cell.* **13:** 445-453.

Martin S.L. (1991) Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol Cell Biol*. **11:** 4804–4807.

Mathias S.L., Scott A.F., Kazazian H.H. Jr, Boeke J.D., Gabriel A. (1991) Reverse transcriptase encoded by a human transposable element. *Science*. **254(5039):** 1808-1810.

McClintock B. (1948) Mutable loci in maize: Nature of the *Ac*action. The mutable *c*loci. The mutable *wx*loci. Conclusions. *Carnegie Institution of Washington.* ***47***:155–169.

McClintock B. (1950) The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci USA.* **36:** 344-355.

McCue A.D., Nuthikattu S., Slotkin R.K. (2013) Genome-wide identification of genes regulated in *trans* by transposable element small interfering RNAs. *RNA Biology*.**10:** 1379–1395.

Miller D.J., Schwartz M.D., Ahlquist P. (2001) Flock house virus RNA replicates on outer mitochondrial membranes in *Drosophila* cells. *J Virol*. **75:** 11664–11676.

Novina C.D., Sharp P.A. (2004) The RNAi revolution. *Nature*. **430(6996):** 161-164.

Pace J.K., Feschotte C. (2007) The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res*. **17:** 422-432.

Parks A.R., Li Z., Shi Q., Owens R.M., Jin M.M., Peters J.E. (2009) Transposition into replicating DNA occurs through interaction with the processivity factor. *Cell*. **138:** 685-695.

Piednoël M., Gonçalves I.R., Higuet D., Bonnivard E. (2011) Eukaryote DIRS1-like retrotransposons: an overview. *BMC Genomics*. **12:** 621.

Platt J.L., Kent N.A., Harwood A.J., Kimmel A.R. (2013) Analysis of chromatin organization by deep sequencing technologies. *Methods Mol Biol*. 983: 173-83.

Polard P., Chandler M. (1995) Bacterial transposases and retroviral integrases. *Mol. Microbiol*. **15:** 13–23.

Raper K.B. (1935) *Dictyostelium discoideum*, a new species of slime mold from decaying forest leaves. *Journal of Agricultural Res.* **50:** 135-147.

Richmond T.J., Davey C.A. (2003) The structure of DNA in the nucleosome core. *Nature*. **423:** 145–150.

Rosen E., Sivertsen A., Firtel R.A. (1983) An unusual transposon encoding heat shock inducible and developmentally regulated transcripts in *Dictyostelium*. *Cell*. **35(1):** 243-251.

Rouget C., Papin C., Boureux A., Meunier A.C., Franco B., Robine N., Lai E.C., Pelisson A., Simonelig M. (2010) Maternal mRNA deadenylation and decay by the piRNA pathway in the early *Drosophila* embryo. *Nature*. **467:** 1128-1132.

Sanchez R., Zhou M.M. (2011) The PHD finger: a versatile epigenome reader. *Trends Biochem. Sci.* **36(7)**: 364-372.

Sandmeyer S.B., Hansen L.J., Chalker D.L. (1990) Integration specificity of retrotransposons and retroviruses. *Annu. Rev. Genet*. **24:** 491-518.

SanMiguel P., Tikhonov A., Jin Y.K., Motchoulskaia N., Zakharov D., Melake-Berhan A., Springer P.S., Edwards K.J., Lee M., Avramova Z., Bennetzen J.L. (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science*. **274:** 765- 768.

Schaap P. (2011) Evolution of developmental cyclic AMP signalling in the *Dictyostelia* from an amoebozoan stress response*. Dev Growth Differ*. **53(4):** 452–462.

Schaap P., Winckler T., Nelson M., Alvarez-Curto E., Elgie B., Hagiwara H., Cavender J., Milano-Curto A., Rozen D.E., Dingermann T., Mutzel R., Baldauf S.L. (2006) Molecular phylogeny and evolution of morphology in the social amoebas. *Science.* **314(5799):** 661-663.

Schilde C., Schaap P. (2013) The *Amoebozoa*. *Methods Mol Biol*. **983:** 1-15.

Schmith A., Groth M., Ratka J. Gatz S., Spaller T., Siol O., Glöckner G., Winckler T. (2013) Conserved gene regulatory function of the carboxy-terminal domain of *dictyostelid* C-module-binding factor. *Eukaryot. Cell*. **12(3)**: 460-468.

Schumann G., Zündorf I., Hofmann J., Marschalek R., Dingermann T.. (1994) Internally located and oppositely oriented polymerase II promoters direct convergent transcription of a LINE-like retroelement, the *Dictyostelium* Repetitive Element, from *Dictyostelium discoideum*. *Mol. Cell. Biol*. **14**: 3074-3084.

Shapiro J.A. (1979) Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. *Proc Natl Acad Sci USA*. **76:** 1933-1937.

Sijen T., Fleenor J., Simmer F., Thijssen K.L., Parrish S., Timmons L., Plasterk R.H., Fire A. (2001) On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell*. **107:** 465–476.

Sijen T., Vijn I., Rebocho A., van Blokland R., Roelofs D., Mol J.N.M., Kooter J.M. (2001) Transcriptional and posttranscriptional gene silencing are mechanistically related. *Current Biology*, **11:** 436–440.

Sillo A., Bloomfield G., Balest A., Balbo A., Pergolizzi B., Peracino B.,Bozzaro S. (2008). Genome-wide transcriptional changes induced by phagocytosis or growth on bacteria in *Dictyostelium*. *BMC Genomics*. **9:** 291.

Siol O., Boutliliss M., Chung T., Glöckner G., Dingermann T., Winckler T. (2006) Role of RNA polymerase III transcription factors in the selection of integration sites by the *Dictyostelium* non-long terminal repeat retrotransposon TRE5-A. *Mol. Cell. Biol.* **26(22)**: 8242-8251.

Siol O., Dingermann T., Winckler T. (2006a) The C-module binding factor mediates expression of the *Dictyostelium* aggregation-specific adenylyl cyclase ACA. *Eukaryotic Cell* **5(4):** 658-664.

Siol O., Spaller T., Schiefner J., Winckler T. (2011) Genetically tagged TRE5-A retrotransposons reveal high amplification rates and authentic target site preference in the *Dictyostelium discoideum* genome. *Nucleic Acids Res.* **39(15)**: 6608-6619.

Siomi H., Siomi M.C. (2009) On the road to reading the RNA interference code. *Nature*. **457(7228):** 396-404.

Stajdohar M., Jeran L., Kokosar J., Blenkus D., Janez T., Kuspa A., Shaulsky G., Zupan B. (2015) dictyExpress: visual analytics of NGS gene expression in *Dictyostelium*. [https://www.dictyexpress.org]

Sucgang R., Chen G., Liu W., Lindsay R., Lu J., Muzny D., Shaulsky G., Loomis W., Gibbs R., Kuspa A. (2003) Sequence and structure of the extrachromosomal palindrome encoding the ribosomal RNA genes in *Dictyostelium*. *Nucleic Acids Res.* **31(9):** 2361-2368.

Szafranski K., Glöckner G., Dingermann T., Dannat K., Noegel A.A., Eichinger L., Rosenthal A., Winckler T. (1999) Non-LTR retrotransposons with unique integration preferences downstream of Dictyostelium discoideum tRNA genes. *Mol Gen Genet*. **262(4-5):** 772-780.

Szak S.T., Pickeral O.K., Makalowski W. Boguski M.S., Landsman D., Boeke J.D. (2002) Molecular archeology of L1 insertions in the human genome. *Genome Biol.* **3(10)**: research0052.

Takeuchi T., Yamazaki Y., Katoh-Fukui Y., Tsuchiya R., Kondo S., Motoyama J., Higashinakagawa T. (1995) Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. *Genes Dev*. **9:** 1211-1222.

Trewick S.C., McLaughlin P.J., Allshire R.C. (2005) Methylation: lost in hydroxylation? *EMBO Reports.* **6:** 315–320.

Tsukada Y., Fang J., Erdjument-Bromage H., Warren M.E., Borchers C.H., Tempst P., Zhang Y. (2006) Histone demethylation by a family of JmjC domain-containing proteins. *Nature*. **439:** 811–816.

Tuschl T., Zamore P.D., Lehmann R., Bartel D.P., Sharp P.A. (1999) Targeted mRNA degradation by double-stranded RNA in vitro. *Genes Dev*. **13:** 3191–3197.

Vaistij F.E., Jones L., Baulcombe D.C. (2002) Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. *Plant Cell*. **14:** 857–867.

van Holde K.E. (1989) Chromatin. *Springer New York*.

van Rij R.P., Saleh M.C., Berry B., Foo C., Houk A., Antoniewski C., Andino R. (2006) The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in *Drosophila melanogaster*. *Genes Dev*. **20(21):** 2985-2995.

Venter P.A., Schneemann A. (2008) Recent insights into the biology and biomedical applications of Flock House virus. *Cell Mol Life Sci*. **65:** 2675–2687.

Vinayachandran V., Pusarla R.H., Bhargava P. (2009) Multiple sequence-directed possibilities provide a pool of nucleosome position choices in different states of activity of a gene. *Epigenetics Chromatin*, **2(1):** 4.

Vogel G., Thilo L., Schwarz H., Steinhart R. (1980) Mechanism of phagocytosis in *Dictyostelium discoideum*: phagocytosis is mediated by different recognition sites as disclosed by mutants with altered phagocytotic properties. *J Cell Biol*. **86:** 456–465.

Voinnet O., Vain P., Angell S., Baulcombe,D.C. (1998) Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. *Cell*. **95:** 177–187.

Wassenegger M., Krczal G. (2006) Nomenclature and functions of RNA-directed RNA polymerases. *Trends Plant Sci*. **11:** 142–151.

Watts D.J., Ashworth J.M. (1970) Growth of myxamoebae of the cellular slime mould *Dictyostelium discoideum* in axenic culture. *Biochem. J.* **119:** 171-174.

Welsh J.D., Leibowitz M.J. (1982) Localization of genes for the double-stranded RNA killer virus of yeast. PNAS. **79:** 786–789.

Westenberger S.J., Cui L., Dharia N., Winzeler E., Cui L. (2009) Genome-wide nucleosome mapping of *Plasmodium falciparum* reveals histone-rich coding and histone-poor intergenic regions and chromatin remodeling of core and subtelomeric genes. *BMC Genomics.* **10**: 610.

Wiegand S., Meier D., Seehafer C., Malicki M., Hofmann P., Schmith A., Winckler T., Földesi B., Boesler B., Nellen W., Reimegard J., Käller M., Hällman J., Emanuelsson O., Avesson L., Söderbom F., Hammann C. (2014) The *Dictyostelium discoideum* RNA-dependent RNA polymerase RrpC silences the centromeric retrotransposon DIRS-1 post-transcriptionally and is required for the spreading of RNA silencing signals. *Nucleic Acids Res*. **42(5):** 3330-3345.

Winckler T. (2013) Selfish DNA: a pharmaceutical perspective. *Pharmazie.* **68:** 467–473.

Winckler T., Dingermann T., Glöckner G. (2002) *Dictyostelium* mobile elements: strategies to amplify in a compact genome. *Cell. Mol. Life Sci*. **59:** 2097–2111.

Winckler T., Iranfar N., Beck P., Jennes I., Siol O., Baik U., Loomis W.F., Dingermann T. (2004) CbfA, the Cmodule DNA-binding factor, plays an essential role in the initiation of *Dictyostelium discoideum* development. *Eukaryotic Cell.* **3(5):** 1349-1358.

Winckler T., Schiefner J., Spaller T., Siol O. (2011) *Dictyostelium* transfer RNA gene-targeting retrotransposons: Studying mobile element-host interactions in a compact genome. *Mob Genet Elements*. **1(2):** 145-150.

Winckler T., Szafranski K., Glöckner G. (2005) Transfer RNA gene-targeted integration: an adaptation of retrotransposable elements to survive in the compact *Dictyostelium discoideum* genome. *Cytogenet Genome Res*. **110:** 288–298.

Winckler T., Trautwein C., Tschepke C., Neuhäuser C., Zündorf I., Beck P., Dingermann T. (2001). Gene function analysis by amber stop codon suppression: CMBF is a nuclear protein that supports growth and development of *Dictyostelium* amoebae. *J. Mol. Biol.* **305:** 703-714.

Windhof I.M., Dubin M.J., Nellen W. (2013) Chromatin organisation of transgenes in *Dictyostelium*. *Pharmazie*. **68:** 595-600.

Xiong Y., Eickbush T.H. (1990) Origin and evolution of retroelements based on their reverse transcriptase sequences. *EMBO J*. **9:** 3353–3362.

Zaratiegiu M. (2013) Influence of long terminal repeat retrotransposons in the genomes of fission yeasts. *Biochem Soc Trans*. **41(6):** 1629-33.

Zuker C., Cappello J., Chisholm R.L., Lodish H.F. (1983) A repetitive *Dictyostelium* gene family that is induced during differentiation and by heat shock. *Cell*. **34(3):** 997-1005.

dictyBase: www.dictybase.org; **Fey P, Gaudet P, Pilcher KE, Franke J, Chisholm RL.** (2006) DictyBase and the Dicty Stock Center. *Methods Mol Biol*. **346:** 51-74.

9. Anhang

9.1 Ergänzungen Manuskript 2

Conserved Gene Regulatory Function of the Carboxy-Terminal Domain of Dictyostelid C-Module-Binding Factor

---

Anika Schmith, Marco Groth, Josephine Ratka, Sara Gatz, Thomas Spaller, Oliver Siol, Gernot Glöckner, Thomas Winckler

Publiziert in: *Eukaryotic Cell*, 2013

Conserved gene-regulatory function of the carboxy-terminal domain of dictyostelid C-module-binding factor

Anika Schmith, Marco Groth, Josephine Ratka, Sara Gatz, Thomas Spaller, Oliver Siol, Gernot Glöckner and Thomas Winckler

Supplemental material

Fig. S1: Correlation between RNA-seq samples.
The heatmap image is based on Spearman correlation of read count data of all genes by comparing all samples against all samples. The color code shows the highest correlation values in green and lowest correlation values in red. The dendrogram reflects the expression-based relation of the samples.

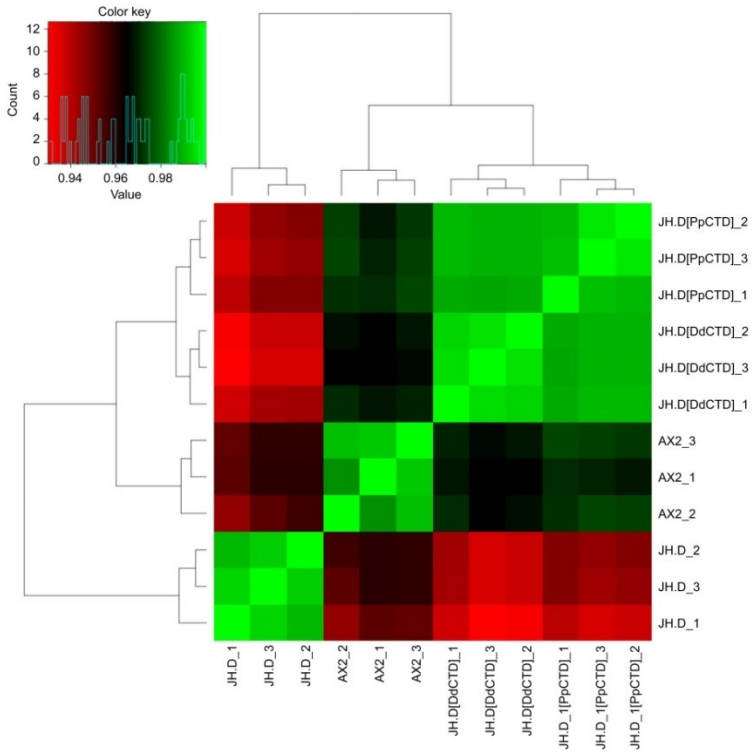


Fig. S2: Calculation of the number of CbfA-regulated genes.
The numbers of 2,298 statistically significant regulated genes were plotted versus the indicated minimal fold changes. The numbers of **CbfA-stimulated** and **CbfA-repressed** genes are indicated.

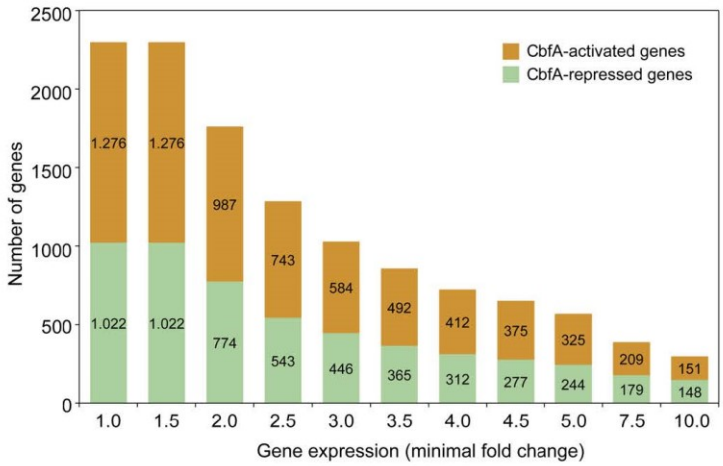
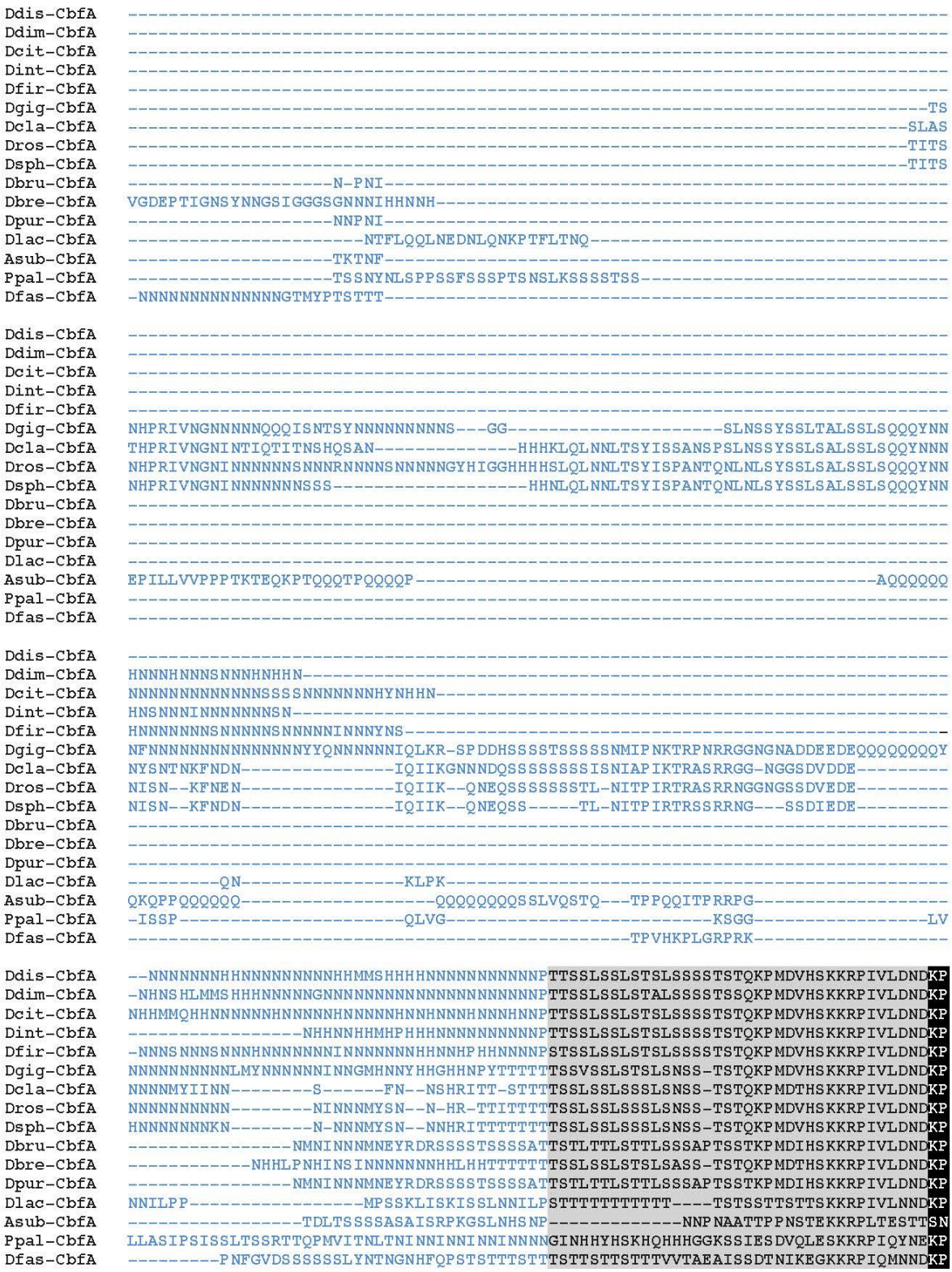


Fig. S3: Multiple sequence alignment of CbfA proteins.
The sequence alignment was generated using COBALT. The JmjC domains are boxed in **yellow color**. The cysteine and histidine residues of zinc finger regions ZF1 and ZF2 are written in **red color**. The asparagine-rich domains are indicated in **blue letters**. The CbfA-CTDs are written in **grey boxes**. The AT-hooks are indicated in **black boxes**.

```
Ddis-CbfA                                                                    MEELIKAPNSNFIIMSN
Dgig-CbfA                                                                    MDDLLKAPNSNFMLVPN
Dpur-CbfA   MNINNINNSSNSSVGNNNNNSNNNININININNNNNNNINNNNNNNNNNNNSNRYSTVEDLQSFNIEELLSAPDTNFILIKS
Dlac-CbfA                                              MQVDTTKLLDHPIFVPKVEDLLDLSFEDKINQPNSNFLSAPV
Asub-CbfA                                              MQIDRSRYLELPIFVPSVESLMSFDIDQLLA--ESNFKMAPT
Ppal-CbfA                                              MQ-TESRYLEYPMYVPPIQSLLSFNIDELLNAPNSNFEMAPK
Dfas-CbfA                                                                    MPNTNFEYAPT

Ddis-CbfA   QPYQTTSPEIVEDYII-KRGQPFVLTGTTQGWSRSNMFT-LD-FLS--------------ERYSEMELINSPRNNETHTD
Ddim-CbfA   -------------------EPPFVLTGTTQGWSRINMFS-LD-FLS--------------ERYSEMELINSPRNNETHTD
Dcit-CbfA   ----------------------------QVQLKVGQEIICLHLNFLS--------------ERYSEMELINSPRNNETHTD
Dint-CbfA   -------------------ESPFVLTGTTQGWSRNNMFT-LD-FLS--------------ERYSEMELINSPRNNETHTD
Dfir-CbfA   ------------SGDKLLKRTPFVLTGTTQGWSRNNMFT-LD-FLA--------------ERYSEMELKNSPRNNETHTD
Dgig-CbfA   QPYQTTSPETVEEYII-KRGHPFVLTGTTNGWSKTSMFS-ID-YLS--------------EKYSEMNLINSPRNSETHTD
Dcla-CbfA   ---------------------PFVLTGTTQEWSKSSMFS-ID-FLS--------------EKYSNMKLHNTPRNSETLQD
Dros-CbfA   -------------------RQPFVLTGTTQGWSKSSMFS-ID-FLN--------------ENYSAMKLNNTPRNSETLQD
Dsph-CbfA   -------------------RSPFVLTGTTQGWSKSSIFS-ID-FLN--------------ENYSTMKLNNTPRNSETLQD
Dpur-CbfA   QSYSSTSPE-TVEESIVKNGLPFVLSDTTHGWSRNQMFS-ID-FLA--------------ERYSGLELINSPRNIEIQKD
Dlac-CbfA   LRYPTITNSIIDDHLIRKGIPLVLSETNKCFTQQTNTLPTPDTKYDHPTKFTLEWLR---DNYSEIKLVNSPTNLDTLTD
Asub-CbfA   LQYTDDISEQTINQYIMERGVPLVLTGTMALWAEQEQHQLHLVNRNDPTQQELFTLEYLSRVFGHIPLHNSPRDNEKHTD
Ppal-CbfA   IPYSSQLGEGAIHQYTITKGVPLVLTNTLEGWDKDLFS-LD-YLH--------------DQFSHIDLIHSPRDNETHTD
Dfas-CbfA   -PYSIDLTESDIQEYTINKGIPLVLSNSMSNWSKDVFS-LD-FLQ--------------ANYSDQKMINSPRNNETHSD
```

```
Ddis-CbfA   LQGWRMKDFISYLQVSPEERNPKHLYGKDIACPREWQEYLSHKLQPQYSYKSR-FDLVSHLPDYLQPETLLVYIGSNGTY
Ddim-CbfA   LQGWRMKDFISYLQVSPEERNPKHLYGKDITCPREWQEYLSHKLQPQYSYKSR-FDLVSHLPDYLQPDTLLVYIGSNGTY
Dcit-CbfA   LQGWRMKDFISYLQVSPEERNPKHLYGKDIQCPREWQEYLSHKLQPQYSYKSR-FDLVSHLPDYLQPDTLLVYIGSNGTY
Dint-CbfA   LQGWRMKDFISYLQVLPEERNPKHLYGKDIACPREWQEYLSHKLQPQYSYKSR-FDLVSHLPDYLQPDTLLVYIGSNGTY
Dfir-CbfA   LQGWRMKDFISYLQVSPEERNPKHLYGKDIACPREMQEYLAHKLQPQYSYKSR-FDLVSHLPDYLQPETLLVYIGSNGTY
Dgig-CbfA   LKGWKMKDFISYLQVSPEERNPKYLYGKDIVCPREWQEYLSNKLQPQYSYKSR-FDLVSHLPEYLQPDTLMCYIGSSGTY
Dcla-CbfA   YQDWTMKEFIQYLQVSPEERNPKYLYGKDLACPIEWQNYLSDNLQPQYSYKSR-FDLVSHLPDYLQPETLMCYIGSNGTY
Dros-CbfA   YQDWTMKEFIQYLQVSPEERNPKYLYGKDLGCPVEWQNYLSDKLQPQYSYKSR-FDLVSHLPDYLQPETLMCYIGSNGTY
Dsph-CbfA   YQDWTMKEFIQYLQVSPEERNPKYLYGKDLACPIEWQNYLSDKLQPQYSYKSR-FDLVSHLPDYLQPETLMCYIGSNGTY
Dpur-CbfA   LQGWRMRDFISYLQVSPEERNPKYLYGKNMTCPREWQDYLSHKLQPQYSYKSR-FDLVSHLPNYLQPDTLMSYIGGGGTY
Dlac-CbfA   EKDWTLGRYLEYLQILPEERNPKHLYAKDVHCPAEWKLLVKEKIQKVFQSQGP-FDLIGNLPQYLRPETSLLYIGSNGTY
Asub-CbfA   CSDWTLGQYIDYLDMIPEERNPKHLYGKDVPCPKEWQNYIETNLTKLLVYRDRRADLIASLPEYLKPETLMLYIGANGTY
Ppal-CbfA   LKGWTVGKFCDYLKVSPEERNPKYLYAKDVQCPTPWRDYLSKKLAKNFVYHSS-YDMMANLPEYLRTDNLQVYIGSNGTY
Dfas-CbfA   LEGWTVGQYINYLQISPEERNPKYLYAKDVLCPKEWADQ-QKKLSKKFVALSS-YDLMSSLPPYLRSENLMFYIGSNGTY

Ddis-CbfA   TPGHIDMCGSLSQNLMVSSDQD--------AFAWWFIVPTEYKDEALKFWGDKGGDVYN-ESRFIRPIDLLGAPFPIYVF
Ddim-CbfA   TPGHIDMCGSLSQNLMVSSDSD--------AHSWWFVVPTEFKDEALKFWGDKGGDVYN-ESRFIRPVDLLGAPFPVYVY
Dcit-CbfA   TPGHIDMCGSLSQNLMVSSDPD--------AYSWWFIVPTEYKDEALKFWRDKGGDVFN-ESRFIRPVDLLGAPFPVYVY
Dint-CbfA   TPGHIDMCGSLSQNLMVSSDPD--------AYSWWFIVPTEYKDEALKFWRDKGGDVFN-ESRFIRPVDLLGAPFPVYVY
Dfir-CbfA   TPGHIDMCGSLSQNLMVSSDQD--------AHSWWFIVPTEYKDEALKFWSDKGGDVYN-ESRFIRPVDLLGAPFPVYVY
Dgig-CbfA   TPGHIDLCGSLGQNLMISNEPD--------ACAWWFIVPTEYKDEAIKFWGDRGGDVYN-ESTFIRPIDLAKAPFPVFVY
Dcla-CbfA   TPGHIDLCGSLGQNLMVSSDTE--------AQSWWFIIPTEYKDEAIKFWGDKGGDVYN-ESSFIKPTELINAPFPVYVH
Dros-CbfA   TPGHIDLCGSLGQNLMVSSDAE--------AHSWWFIIPTEYKDEAIKFWGDKGGDVYN-KSSFIKPVELLNAPFPVYIY
Dsph-CbfA   TPGHIDLCGSLGQNLMVSSDTD--------AHAWWFIIPTEFKDEAIKFWGDKGGDVYN-ESSFIKPAELLNAPFPVYIY
Dpur-CbfA   TPGHVDLCGSLGQNLMVAAEGD--------SFAYWFIVPSRYRETAIKFWNDRGGDIYSGDSHFINPIELLGAPFPVYVY
Dlac-CbfA   TNGHVDIAGSNGHNIMVSAEPD--------SFSYWFIVPRDYKQEAIKFWAELGGDIFN-ESRFLKPSDLLKASFPIYIV
Asub-CbfA   TAGHMDNVGTRGHNLMLHARPEETDVNADNTFAYWFVLETKSMSKAIEFWSDKGGDIFN-ESRFLNPKEFIGAPFDVYVI
Ppal-CbfA   TPGHMDNVGANGHNIMVGSQDS------DSSSAYWFVVSRQDADAAIEFWTSKGGDIFN-ESRFLNPRELLHAPFKVYVI
Dfas-CbfA   TPGHVDLVGSNGHNIMVHSDPN--------AYALWLVVGSKDRDKAFKFWNEKGGDIFN-ENLFVNPRDLADAPFKVYVI

Ddis-CbfA   KQRPGDFIFVPPDSVHQVVNCGPGISTKVAWNSISLKSLPISYFSSLPHTRRMAKPELFRIKAIAYYTLRKIMGDVENTN
Ddim-CbfA   KQRPGDFIFVPPDSVHQVINCGPGISTKVAWNSISLKSLPISYFSSLPHTRRMAKPELFRIKAIAYYTLRKIMGDVENTN
Dcit-CbfA   KQRPGDFIFVPPDSVHQVLNCGPGISTKIAWNSISLKSLPISYFSSLPHTRRMAKPELFRIKAIAYYTLRKIMGDVENTN
Dint-CbfA   KQRPGDFIFVPPDSVHQVLNCGPGISTKIAWNSISLKSLPISYFSSLPHTRRMAKPELFRIKAIAYYTLRKIMGDVENTN
Dfir-CbfA   KQRPGDFIFVPPDSVHQVVNCGAGISTKIAWNSISLKSLPISYFSSLPHTRRMAKPELFRIKAIAYYTLRKIMGDVENTN
Dgig-CbfA   KQRPGDFMFIPPDSVHQVINVGSGISIKAAWNSISLKSLPISYFSSLPQARRMAKPELFRIKAVAYYTLRKIMGDVESTN
Dcla-CbfA   KQRSGDLVFIPPDSVHQVINVGTGTSIKCAWNSISLKSLPISYFSSLPHTRRLAKPELYRIKAVAYYTLRKIMCDVENTN
Dros-CbfA   KQSAGDLVFIPPDSVHQVINVGSGISIKCAWNSISLKSLPISYFSSLPHTRRLAKPELYRIKAVAYYTLRKIMCDVENTN
Dsph-CbfA   KQSAGDLVFIPPDSVHQVINVGSGTSIKCAWNSISLKSLPISYFSSLPHTRRLAKPELYRIKAIAYYTLRKIMCDVENTN
Dpur-CbfA   RQRIGDFVFIPPDSLHQVMNFG-GTSIKLAWNSISLKSLPISYYSSLPISKKLAKPEMFRIKAVAYYTLRKIMGDVENTN
Dlac-CbfA   KQGLGDFVLVPPDAVHQVINHG-GTTLKMAWNTITIKSLPISYFHSLPDYKQVSKPEVYRIKAVAYYSLKKMIQQIEVEN
Asub-CbfA   RQRVGDLVIIPPSAVHQVINIG-GPTIKVAWNTLPVSCAPISYYTMLPLQRSISKPEIYRVKATIYYSLKKIMEKLADGL
Ppal-CbfA   RQQVGDFVLIPPESVHQVINIG-APSIKIAWNTVSARTLPISYYSAMPRQRKHLKPELYRIKAIVYYTLRKKMSVLQEM-
Dfas-CbfA   KQKIGDLVLIPPESVHQVMNVG-GVTVKMAWNTISTASLPVAYKSALPLARKFGKPETYSIKAIIYYTLKKKIASLQEL-

Ddis-CbfA   ----------------------FNTIDVNDVIDIIAPLLEIFHNILQTESILIPKPNYPYC-NGETIPFLQPFKYFNGDR
Ddim-CbfA   ----------------------FNTIDVNDVIDIIAPLLEIFHNILQTESILIPKPNYPYC-NGETIPFLQPFKYFNGDR
Dcit-CbfA   ----------------------FNTIDVNDVIDIIAPLLEIFHNILQTESILIPKPNYPYC-NGETIPFLQPFKYFNGDR
Dint-CbfA   ----------------------FNAIDVNDVIDIIAPLLEIFHNILQTESILIPKPNYPYC-NGETIPFLQPFKYFNGDR
Dfir-CbfA   ----------------------FNAIDVNDVIDIIAPLLEIFHNILQTESILIPKPNYPYC-NGETIPFLQPFKYFNSDR
Dgig-CbfA   ----------------------FMTIDVNDVIDIISPLLEIFHNILQTESILIPKPNYPYC-NGETIPFLQPFKYFTGDR
Dcla-CbfA   ----------------------FMTIDVNDVIDIISPLLEIFHNILQTESILIPKPNYPYC-NGETIPFLQPFKYFGGDR
Dros-CbfA   ----------------------FMTIDVNDVIDIISPLLEIFHNILQTESILIPKPNYPYC-NGETIPFLQPFKYFGGDR
Dsph-CbfA   ----------------------FMTIDVNDVIDIISPLLEIFHNILQTESILIPKPNYPYC-NGETIPFLQPFKYFGGDR
Dpur-CbfA   ----------------------FSTIDVNDVIDIISPLLEIFHNILQVESILIPKPNYPYC-NGETIPFLQPFKYFGGER
Dlac-CbfA   ------------------------PVDIKTTIENLSSMIEIFHHILQTESILIPKPNYPFCGDGESIPFLKPFKYFGTEK
Asub-CbfA   AGNEEGKL---------MN-------------EYVEPLLGILHHILQSESILIPKPSYPFAAPGETIPFLKPFKLITTEL
Ppal-CbfA   --NLNC--------------------DLPATLEYITPLLEIFHHFVQSESILIPKPSYPYAARDETIPFLKPFKLFASQF
Dfas-CbfA   ---TNY--------------------DSQSIVDYLSPLLDLYSHMIRSESILIPKPPYPFCPHNEVITFLNPFKNFSSPF

Ddis-CbfA   IQDRRCDHCNSDIFNRCYHCETCKT--DDGQGKDFCFDCVSSGIGCEFHFKVMVLKEFISHSKLKKELSSFYEIYKNL-L
Ddim-CbfA   IQDRRCDHCNSDIFNRCYHCETCKT--EDGQGKDFCFDCVSSGIGCEFHFKVMVLKEFISHSKLKKELSSFYEIYKNL-L
Dcit-CbfA   IQDRRCDHCNSDIFNRCYHvETCKT--DDGQGKDFCFDCVSSGIGCEFHFKVMVLKEFISHSKLKKELSSFYEIYKNL-L
Dint-CbfA   IQDRRCDHCNSDIFNRCYHCETCKT--DDGQGKDFCFDCVSSGIGCEFHFKVMVLKEFISHSKLKKELSSFYEIYKNL-L
Dfir-CbfA   IQDRRCDHCNSDIFNRCYHCETCKT--EDGQGKDFCFDCVSSGIGCEFHFKVMVLKEFISHSKLKKELSSFYEIYKNL-L
Dgig-CbfA   IQDRRCDHCNSDIFNRCYHCENCKT--DDGYGKDYCFDCVSSGIGCEFHFKSMVLKEFISHSKLKKELGSFYEIYKNL-L
Dcla-CbfA   IQDRRCDHCNSDIFNRCYHCETCKT--EDGQGKDYCFDCVSSGIGCEFHFKTMVLKEFISHSKLKKELSSFYEIYKNL-L
Dros-CbfA   IQDRRCDHCNSDIFNRCYHCETCKT--EDGKGKDYCFDCVSSGIGCEFHFKIMVLKEFISHSKLKKELSSFYEIYKNL-L
Dsph-CbfA   IQDRRCDHCNSDIFNRCYHCETCKT--EDGQGKDYCFDCVSSGIGCEFHFKIMVLKEFISHSKLKKELSSFYEIYKNL-L
Dbru-CbfA                 HRCYHCDQCKT--EDGHGKDFCFDCVSMGVGCEFHFKEMVLKEFISHSKLKKELSNFYDIYKNL-L
Dbre-CbfA                   CYHCETCKT--QNGNGKDYCFDCVSSGIGCEFHFKVMVLKEFISHSKLKKELSSFYEIYKNL-L
Dpur-CbfA   IQDRRCEHCHSDIFNRCYHCDQCKT--DDGHGKDFCFDCVSMGVGCEFHFKEMVLKEFISHSKLKKELSNFYDIYKNL-L
Dlac-CbfA   IQDRRCDHCNCDIFNRCYHCQDCKSK-EDPFGKDYCFDCVAKGFGCEFHFKDMILKEYISHSKLKKELVKYYQVFKNL-L
Asub-CbfA   KHDRKCDQCNCDIFNRCYTCPLCES--------RDFCFDCVSHGNGCEDHFAQMELREFISHHKMRKDLDTYYDIYKAYLL
Ppal-CbfA   VHDRRCDHCNSDIFNRCYHCPVCKNNDDEVNGRDYCFECVSRGFGCEYHFSEMSLKEYISHSKIKKELDGFYEIYSEL-M
Dfas-CbfA   LHDRKCDHCNCNIFNRAYHCQLCAD--KENNGKDYCFDCVSRGYGCELHFQSMSLKEFISHSKLKKELETFFDKYTTL-L
```


Ddis-CbfA AHSGRRPKE---VDDIITKSTDRVSDECGFLTTATVAYHVVFYSSQKKIKCHRCEKRFKKFSIIFCTNCNARFCEQCVVN
Ddim-CbfA AHSGRRPKE---VDDIITKSTDRVSDECGFLTTATVAYHVVFYSSQKKIKCHRCEKRFKKFSIIFCTNCNSRFCEQCVVN
Dcit-CbfA AHSGRRPKE---VDDIITKSTDRVSDECGFLTTATVAYHVVFYSSQKKIKCHRCEKRFKKFSIIFCTNCNSRFCEQCVVN
Dint-CbfA AHSGRRPKE---VDDIITKSTDRVSDECGFLTTATVAYHVVFYSSQKKIKCHRCEKRFKKFSIIFCTNCNSRFCEQCVVN
Dfir-CbfA AHSGRRPKE---VDDIITKSTDRVSDDCGFLTTATVAYHVVFYSSQKKIKCHRCEKRFKKFSIIFCTNCNSRFCEQCVVN
Dgig-CbfA NHSGRRPKE---VEDIITKTTERVSEECGFITTATVAYHVVYYISQKKIKCHRCEKRFKKFSIIFCTQCSSRFCEQCVVN
Dcla-CbfA THSGRRPKE---VEDIITKTTDRVSEDCGFLTTATVAYHVVYYSSQKKIKCHRCEKRFKKFSIIFCTQCSSRFCEQCVVN
Dros-CbfA THSGRRPKE---VDEIITKTTDHVSKDCGFLTTATVAYHVVYYSSQKKIKCHRCEKRFKKFSIIFCTQCSSRFCEQCVVN
Dsph-CbfA THSGRRPKE---VDEIITKTTDRVSEDCGFLTTATVAYHVVYYSSQKKIKCHRCEKRFKKFSIIFCTQCSSRFCEQCVVN
Dbru-CbfA LHSGRRQRD---VDEFIAKTSELVVEECGFLTTATVAYHVVYYSSQKKIKCHRCERRYKKFSVVFCTNCSSRFCEQCVIG
Dbre-CbfA MHSGRHSKE---VEEIITKSTERVTEECGFLTTATVAYHVVYYSSQKKIKCHRCEKRFKKFSIIFCTACSSRFCEQCVIN
Dpur-CbfA LHSGRRQRD---VDEFIAKTSELVVEECGFLTTATVAYHVVYYSSQKKIKCHRvERRYKKFSVVFCTNCSSRFCEQCVIG
Dlac-CbfA AQTGMKEQN---IEDLVQSRLNEMLDRCKFLTTATVAYHVVYFSGQDEQSCHRCQKFYKKYNLVHCTNCKSRFCEKCVCQ
Asub-CbfA AQTKQKKQDHEDIDTLITNKQNAIVQKCGFLTTATIAYHVVYHANKKE-KCHQCTTLLPKFSVVYCTHCRKPYCDRCVAG
Ppal-CbfA LRSGNKRKK---VDEVIQEKQDKIFERCGFLTTATLAYHAVYFSTQNEVKCHQCERRYKKHSIVFCSHCPARYCKSCCQS
Dfas-CbfA INTGIKPKE---ANDIVEKKKALIDTQCGYLTPATIAYHVVFYSNSTLTKCHHCGDAKQKHSVVYCCSCPARYCESCVAK
Ddis-CbfA TFGQNFQVLMKRNEWECFCCKGLCDCSNCTSNSNSSNHPRILNNNQQLGLPYNNNNNSNNNNNNNINNNNNNNNNNINNN
Ddim-CbfA TFGQNFQVLMKRNEWECFCCKGLCDCSNCSSNSNSLNHQRIVNGNQQLSLPYNNNNNNNNNNNNNNNNNNNNNNNNNNN
Dcit-CbfA TFGQNFQVLMKRNEWECFCCKGLCDCSNCNSN-NSSNHPRIVNGNQQLNLPYNNNNNNNNINNNNNINNNNNSNNNNNNN
Dint-CbfA TFGQNFQVLMKRNEWECFCCKGLCDCSNCNSNSNSSNHPRIVNGNQQLSLPYNNNNNNNNINNNNG--NNNNSNNNNNNN
Dfir-CbfA TFGQNFQVLMKRNEWECFCCKGLCDCSNCNSNPNSSNHPRIVNGNQQLSLPYNNNN----------INNNSNSNNNNNNN
Dgig-CbfA TFGQNFQVLMRRNEWDCYCCKNICDCSNCHNN----------N
Dcla-CbfA NFGQNFQVLMRTNQWDCFCCRNECDCSNCANN----------N
Dros-CbfA TFGQNFQVLMRRNEWECFSCTNQCDCLNCVNN----------
Dsph-CbfA TFGQNFQVLMRRNEWECFCCTNECDCLNCANN----------N
Dbru-CbfA -FKDAFTSIMKRSEWNCYSCQGRCDCTACKNN----------EDNRNLLNN
Dbre-CbfA TFGQNFQILMKRNEWDCYCCQNVCDCSNCLNN----------NNNNNNNNN
Dpur-CbfA -FKDAFTSIMKRSEWNCYSCQGRCDCTACKNN----------EDNRNFLNN
Dlac-CbfA -SGNEFCELIREQEWQCYSCKGTCDCPPCQFN----------K
Asub-CbfA -YNVDFFQIVSSGEWKCFSCLGTCLCTPCVTG--------VLKNDNELKAA----------DSKSNSNSITG
Ppal-CbfA -YSIDFYELISCSNWKCFNCIGICSCIHCSTN----------NNN
Dfas-CbfA IYRVDFFSIVSRDQWKCFRCFDICTCNNCKNPSAIGNLVAGGNNNNNNGGGGGSPTHHHHSNQMSPSLQVAHHQNTVNYS
Ddis-CbfA NNNMNNNN-SINNNNNNNNNNNINNNNINNNNHHNNNG----------NNNLNSSYSSLNALSS
Ddim-CbfA NNNNNNNNNNNNNNNNNNNNSNNNNNINNNNINNNNTINNNNNNNNSIINSNNINNN----NGNNNLNSSYSSLNALSS
Dcit-CbfA NNNNNNNNNNNINNINNINNNNNNNNN-NNNIINNNN---NNNHNNKAIIIKTTILIITIIVNGNNSLNSSYSSLNALSS
Dint-CbfA NNNNNNNNNNN--------NNNNNNNNN-NNN--NNNN---NNN----------NGNN-LNSSYSSLNALSS
Dfir-CbfA NSNNNNNHNNS----------NINNNI----------GNNNLNSSYSSLNALSS
Dgig-CbfA NTSNNNNNNT----------
Dcla-CbfA ISNGNNNNNNNINNPN----------
Dros-CbfA VPN----------T----------
Dsph-CbfA VSN----------T----------
Dbru-CbfA DQDYNSKNNNTLSKYNNNNNNNNNNNNNNNQSPHHQHSPHHQQQSSHNDM---------DGPFKKNSPNLQSLSALPSLSS
Dbre-CbfA NNNNNNHNHNHIQNGNGSNIQLVHSNNNNNNRQIQDSFSNQNQNNNNN----------NNNIPLTSSYSSLSALSS
Dpur-CbfA DQDYNSKNNNTLSKYNNN----------HHQHSPHHQQQSSHNDM---------DGPFKKNSPNLQSLSALPSLSS
Dlac-CbfA LNNNNHNHNNHINQQNSIINSSNGNNNNNHHLLNNNTNNNNNNNNNNIH----------LNSPSSTLSSLSTLSTLST
Asub-CbfA NTSNGIFSNRTFHNINNLMSISNNNNSEQDNSHHENMAHKIRSFSMRRCKVNPPPFA---------PPLEEYNTSLQYQHV
Ppal-CbfA INNSNSINNNNSSKIN-------NNNNSNEYSFSEDS----------KQMSAIKL
Dfas-CbfA NTNNNLHNLNNINNSNFQQHQHNNQNNNNNN---GNDGLPQNQYNHHQH----------SMSGDES
Ddis-CbfA LSQQQSYGSYDNYNNN---NNNNNYNNNNNNNGHIQILKSGRQYDDEQSSSSGSGSSNSTPTKPRPRNGGDDGLMSHFSG
Ddim-CbfA LSQQQQPYGNYDNYNN---NNNYN-----NNNNHIQVLKSGRQYDDEQSSSSGSGSSHSTPTKSRPRNGGDDGLISHFSG
Dcit-CbfA VSQQQS-YGNYDNYNN---NNNNNYNNSNSNNGHIQILKSGRQYDDEQSSSSGSGSSNSTPMKSRPRNGGDDGLMSHFSG
Dint-CbfA LSQQQQPYSNYDNNYN---SNNNN-NNNNYNNNHIQILKSGRQYDDEQSSSSGSGSSNSTPMKSRPRNGGDDGLMSHFSG
Dfir-CbfA LSQQQPYGGNYDNNYNTNTNNNNN-NYNNNNNNNHIQILKSGRQYDDEQSSSSGSGSSNSTPIKSRPRNGGDDGLMSHFSG
Dgig-CbfA ----------
Dcla-CbfA ----------
Dros-CbfA ----------
Dsph-CbfA ----------
Dbru-CbfA ISSSL----------NNNGN--YNNYNNN----------
Dbre-CbfA LSQQQ-----YDVNNN---NNNDN--GNGYNNG-------GGGGFNSISSMIGSNIESTLP-KLRTRRGINGG-----SG
Dpur-CbfA ISSSL----------NNNGN--YNNYNNN----------
Dlac-CbfA LSSLPY----------LSTSNDR----------QLIK----------
Asub-CbfA LAYQPS-------TFTP------N----------IQLLT----------PTTPPLKA----------
Ppal-CbfA LSEQCM-------LTEPI----------PISSSSPPVSSPP----------
Dfas-CbfA LEYQHHQH-QHQQQHN---NNNNN------NNHHHHIFSQTNDFIDR----------


Ddis-CbfA
Ddim-CbfA
Dcit-CbfA
Dint-CbfA
Dfir-CbfA
Dgig-CbfA
Dcla-CbfA
Dros-CbfA
Dsph-CbfA
Dbru-CbfA
Dbre-CbfA
Dpur-CbfA
Dlac-CbfA
Asub-CbfA
Ppal-CbfA
Dfas-CbfA

```
Ddis-CbfA   KGRPPKNLKEWTS-THKFIISLIELFRSSNNAILGKPNPH-------YK--PIEN---LPPLVQLYLSQRKAFGGVLWAK
Ddim-CbfA   KGRPPKNLKEWTS-THKFIISLIELFRSSNNAILGKPNPH-------YK--PIEN---LPPLVQLYLSQRKAFGGVLWAK
Dcit-CbfA   KGRPPKNLKEWTS-THKFIISLIELFRSSNNAILGKPNPH-------YK--SIEN---LPPLVQLYLSQRKSFGGVLWAK
Dint-CbfA   KGRPPKNLKEWTS-THKFIISLIELFRSSNNAILGKPNPH-------YK--PLEN---LPPLVQLYLSQRKAFGGVLWAK
Dfir-CbfA   KGRPPKNLKEWTS-THKFIISLIELFRSSNNAILGKPNPH-------YK--PIEN---LPPLVQLYLSQRKAFGGVLWAK
Dgig-CbfA   KGRPPKNLKEWTS-THKFIISLIELVRSSNNAILGKPNPH-------YK--PIEN---LPPLVQLYLSNRKAFGGVLWAK
Dcla-CbfA   KGRPPKNLKEWTS-THKFIISLIELVRSSNNVILGKPNPH-------YK--PIEN---LPPLVQLYLSQRKSFGGVLWAK
Dros-CbfA   KGRPPKNLKEWTT-THKFIISLIELVRSSNNVILGKPNPH-------YK--PIEN---LPPLVQLYLSQRKCFGGVLWAK
Dsph-CbfA   KGRPPKNLKEWTS-THKFIISLIELVRSSNNVTLGKPNPH-------YK--PIEN---LPPLVQLYLSQRKCFGGVLWAK
Dbru-CbfA   KGRPPKNLKEWTS-THKFIISLVELNRS--NSVLGKNNPN-------YK--RFED---LPTLVQLYLTKRRAFGGVLWAK
Dbre-CbfA   KGRPPKNLKEWTS-THKFIISLIELVRSSNNAILGKPNPN-------YK--PIPS---LPPLVQLYLSSRKSFGGVLWAK
Dpur-CbfA   KGRPPKNLKEWTS-THKFIISLVELNRS--NSVLGKNNPN-------YK--RFED---LPTLVQLYLTKRRAFGGVLWAK
Dlac-CbfA   KGRPAKNLRAWTN-SHKHIIQLIENLRS-NDKSLNNNNGSVKCGTTSTNGATFKPLDQLPELVSLYIKKKQAFGGVMWAK
Asub-CbfA   KGRPPKILKETTSDTHKSVCAILDALRS-----------HIP-----YPTHAVTAPDPLPTAVRLYVNHKDKLGGLLWAK
Ppal-CbfA   KGRPAKNLKDWTD-SHKFILGLIEIYRN------GSKYSS-------TKALPVDT--AVPELVQLYLENRECFGGLLWAK
Dfas-CbfA   KGRPPKNLKEWTS-THKFFINLIESSRK------GTPYHSL----------HINTHEVLPSLISLYIDRKEKFGGVMWAK

Ddis-CbfA   TNSCPLLPCIWVKDLSVIPPNTKLLPSLIQGKKIVVLFFG-DQDQEEYVGIVGKKSIFSFDEVNQTLLLKCGEVPLAQLE
Ddim-CbfA   TNSCPLLPCIWVKDLSVIPPNTK---------------YS-H
Dcit-CbfA   TNSCPLLPCIWVKDLSVIPPNTK---------------YS-H
Dint-CbfA   TNSCPLLPCIWVKDLSVIPPNTK---------------YS-I
Dfir-CbfA   TNSCPLLPCVWVKDLSVIPPNTKLLPSLIQGKKIVVLFFG-D
Dgig-CbfA   TNSCPLLPCLWVKDLSVIPPNTKLLPSLIQGKKIVVLFFG-DLDQEEYVGIVGKKSIFSFDEVNQTLLLKCNETPLLQLE
Dcla-CbfA   TNSCPLLPCIWVKDLSVIPPNTKLLPSLIQGKKIVVLFFG-E
Dros-CbfA   TNSCPLLPCIWVKDLSVIPPNTKLLPSLIQGKKIVVLFFG-ELDQDEYVGIVGKKSIFSFDEVNQTLLLKCNETPLLQLE
Dsph-CbfA   TNSCPLLPCIWVKDLSVIPPNTKLLPSLIQGKKIVVLFFG-ELEQDEYVGIVGKKSIFSFDEVNQTLLLKCNETPLLQLE
Dbru-CbfA   TNSCPLLPCLWVKDLSIIPPNFKLLPSLVQGKKIVVLFFGAT
Dbre-CbfA   TNSCPLLPCLWVKDLTVIPPTTKLLPSLIQGKKIVVLFLV-NV
Dpur-CbfA   TNSCPLLPCLWVKDLSIIPPNFKLLPSLVQGKKIVVLFFGCNIEDEESIGMVGKKSIFSFDEVNQTLLLKCNEVPLKQLE
Dlac-CbfA   TKQIPLLPCLWVKDLSVIPPGTKLLPSLVQGKKILVLFFT--DNTDDRVAMVGKKSIFTFDEVNQTLLLKCPELPLSQLE
Asub-CbfA   TVSCPWMPCVWVKDLHVVPSSLKLLPSLA-GKKIIVVFFG----QPDDFRIVGKKSTMTFVEATETNLVKCGEEVLQQLD
Ppal-CbfA   TINFPLLPCLWVRDLTVIPPSTKLLPSLRVGKKIVVLFFG---NQDEYVGIVGKKSLFAFDEVNQTLLLKCSSEPLRQLE
Dfas-CbfA   TINHPLLPCLWVKDLTVIPPATKLLPSLCNGKKIVVLFFG---DQEEYVGIVGKKSIFSFDEVNQTLLLKCSEGPLKQLE

Ddis-CbfA   DLFNTTEPEIAMKKDIAAFNYKNQIEEKEEGLYVKQELYNNKKII
Dgig-CbfA   DLFSTTEPEIAIKKEIASFNYKNQIEEK
Dros-CbfA   DLFNTTEPEIAIKKEIASFNYKNQMKK
Dsph-CbfA   DLFNTTEPEIAIKKEIASFNYK
Dpur-CbfA   DLFNTVEPDIATKKDISTFNYKTQMEEKEDGLYVKQELYNNNMNMNMNMNNINNINNNNNINNNSNNKYD
Dlac-CbfA   DLFNSADPELATKKEISSFDYRSQIEEKEDGLFIKDSLLINMNSSSSTNNNNNNNGNNNIIGSTINNGNGLDNDIIMSGN
Asub-CbfA   LLFNICTPTVATKVEFNEFNYQSQIVETEEGALVAGTPGNGNRMMVIPTITPTPSPATTTPVAHQS-PKLASSAPVVSDV
Ppal-CbfA   ILFSVTDPLIANKKEIQTFDYKSQIDETEEGLAVKASLYSSTSKLISTTTTTTSTSSTTTTTSLQSFPMLSTSTSPSHSH
Dfas-CbfA   LLFNMTEPVIASKLEIATFNYKNQISESEEGLTIINKPDSESPAMQLANLHNTIAPHPHPHQHQQHHPQQNQHQQQQQTD

Dlac-CbfA   EDLLNGMAALQQLSDQSSIEILQQQQQHQQLQQLQQQHLQQQQPHLPQLPQQHQLQQFQQQTPNVIGNNNLQQII
Asub-CbfA   AATAVATTDTKMQAVPSPASISTYTPLDTTISAALHVLEQARLSTTTTTTTTTTTATKHTAPMVIQRSDVHHPS
Ppal-CbfA   SHSPSSSNSSPTNSNNNIILTPITTSPSSSLMNNNGIKSNIILPSLFPSASSQSSTSTSPPQHTSTKSATSKPSSTSNHL
Dfas-CbfA   YPIVNGTSPT---TSPTLMARMGATNQMPSTPPSKLSSINSLINSTLNNFNNIGNNNNNNNNNNNNNNNNNDINNNTTN

Ppal-CbfA   ANLTSLGSDDHSPPLPPTKHHKSHHQKQPCIHSNLYNNNNNDNNNNNNNNNNNNNNTTLPHIIEKSLSSSSSSTTSPSSPL
Dfas-CbfA   DNGNDNNNNNNNNIQQYQQQQQQQKQSQFQPQIDPNYGLVFNPFNNNDNDNDNSGSDNNNNNNILQVNSYSQHREKSSGIL

Ppal-CbfA   SSSTSTSSLSPTSTSTSASPTQPIIMRK
Dfas-CbfA   PPFFNKQQQQQQDTDHHHHHHSYDQHEDDKESPQQSIADHDDHQNEDPEDGSYPQLKMDLPNS
```

Fig. S4: Zinc finger regions of CbfA and CbfA-like proteins.
The ZF1 and ZF2 regions of dictyostelid CbfA and CbfB proteins were aligned with representative fungal CbfA-like proteins from the following species: *Aspergillus nidulans* (GI:67528200), *Chaetonium globosum* (GI:116199707), *Coccidioides immitis* (GI:119188365), *Gibberella zeae* (GI:46138433), *Magnaporthe grisea* (GI:39973571), *Neosartorya fischeri* (GI:119501326), *Neurospora crassa* (GI:85078924), and *Phaeosphaeria nodorum* (GI:111066574). Note that the *N. crassa* protein is identical to DMM-1. The ZF2 region was also aligned with the zinc-finger domain of *Arabidopsis* monoamine-oxidase A repressor R1 (GI:145360248) and a related zinc finger region in human cell division cycle-associated 7-like protein isoform 2 (GI:188497635).

ZF1

```
Ddis-CbfA       DRRCDHCNSDIFNRCYHCETCKT-(06)-DFCFDCVSSGIGCEFHFKVMV
Dgig-CbfA       DRRCDHCNSDIFNRCYHCENCKT-(06)-DYCFDCVSSGIGCEFHFKSMV
Dros-CbfA       DRRCDHCNSDIFNRCYHCETCKT-(06)-DYCFDCVSSGIGCEFHFKIMV
Dsph-CbfA       DRRCDHCNSDIFNRCYHCETCKT-(06)-DYCFDCVSSGIGCEFHFKIMV
Dpur-CbfA       DRRCEHCHSDIFNRCYHCDQCKT-(06)-DFCFDCVSMGVGCEFHFKEMV
Dlac-CbfA       DRRCDHCNCDIFNRCYHCQDCKS-(07)-DYCFDCVAKGFGCEFHFKDMI
Ppal-CbfA       DRRCDHCNSDIFNRCYHCPVCKN-(07)-DYCFECVSRGFGCEYHFSEMS
Asub-CbfA       DRKCDQCNCDIFNRCYTCPLCES-(01)-DFCFDCVSHGNGCEDHFAQME
Dfas-CbfA       DRKCDHCNCNIFNRAYHCQLCAD-(06)-DYCFDCVSRGYGCELHFQSMS
Ddis-CbfB       KKICIHCNGFIFNRAYLCRECQK-(07)-LICIHCISEG-ACKHKSQILM
Dpur-CbfB       KKICIYCNGFIFNRAYLCSECQK-(07)-LVCIQCISEG-ACKHKNQTLL
Dlac-CbfB       KKVCSYCSAFIFNRAYSCQDCSK-(02)-MICINCIAEG-QCKHKSPTLY
Ppal-CbfB       LLHCADCSTVIFNRSYQHNNQAK-(15)-DYCIYCIAEGKEQQHANDINL
Asub-CbfB       NRFCSGCHAYIFNRCYRAQQMQN-(01)-HFCQECVAEGKDCGLINELEL
Dfas-CbfB       RRTCFGCNAYIFNHFYSCTECYA-(25)-DLCLNCIAEATICKHAKKVQL
Aspergillus     NITCAYCRGNIFNRFLTCPSCAT-(09)-DICMECYVMGRSCACISELKW
Neosartorya     NITCSYCRCNIFNRFLTCPWCAG-(10)-DICMECYVMGRSCACLSRLKW
Coccidioides    NIICSYCRCNIFNRFLTCPSCVG-(10)-DICMECYSIGRSCACISKLKW
Phaeosphaeria   NVTCAYCRCNIFNRFLSCKSCKN-(10)-DVCMDCYCMGRSCGCQSGYTW
Neurospora      CITCSYCRANIFNRFLTCKHCVR-(10)-DICMECYAMGRSCLCVSNLTW
Chaetonium      NITCSYCRANIFNRFLTCKRCVR-(10)-DICMECYVMGRSCFCVSNLSW
Gibberella      CITCSYCRSNIFNRFLTCKHCTR-(10)-DVCMECYAMGRSCACLSGLQW
Magnaporthe     NITCSYCRCNIFNRFLTCKHCIR-(10)-DICMECYAMGRSCVCVSKLSW
Naegleria       RRYCNHCKGDIFNRCFHAEEFED-(02)-DLCMDCVAEGRGYLYKDSLQL
```

ZF2

```
Ddis-CbfA       KIKCHRCEKRFKKFSIIFCTN------CNA---RFCEQCVVNTFGQNFQVLMKRNEWECFCCKGLCDCSNCTSN
Dgig-CbfA       KIKCHRCEKRFKKFSIIFCTQ------CSS---RFCEQCVVNTFGQNFQVLMRRNEWDCYCCKNICDCSNCHNN
Dros-CbfA       KIKCHRCEKRFKKFSIIFCTQ------CSS---RFCEQCVVNTFGQNFQVLMRRNEWECFSCTNQCDCLNCVNN
Dsph-CbfA       KIKCHRCEKRFKKFSIIFCTQ------CSS---RFCEQCVVNTFGQNFQVLMRRNEWECFCCTNECDCLNCANN
Dpur-CbfA       KIKCHRCERRYKKFSVVFCTN------CSS---RFCEQCVI-GFKDAFTSIMKRSEWNCYSCQGRCDCTACKNN
Dlac-CbfA       EQSCHRCQKFYKKYNLVHCTN------CKS---RFCEKCVC-QSGNEFCELIREQEWQCYSCKGTCDCPPCQFN
Ppal-CbfA       EVKCHQCERRYKKHSIVFCSH------CPA---RYCKSCQ-SYSIDFYELISCSNWKCFNCIGICSCIHCSTN
Asub-CbfA       KEKCHQCTTLLPKFSVVYCTH------CRK---PYCDRCVA-GYNVDFFQIVSSGEWKCFSCLGTCLCTPCVTG
Dfas-CbfA       LTKCHHCGDAKQKHSVVYCCS------CPA---RYCESCVAKIYRVDFFSIVSRDQWKCFRCFDICTCNNCKNP
Ddis-CbfB       KATCHQCKLARPRNKIAYCCH-(16)-CQK---KFCSQCLWNRYQVQLVDCLSQRKWECLFCLNKCNCSACNRK
Dpur-CbfB       KATCHQCKLARPRNKIAYC-H-(16)-CQK---KFCSQCLWNRYQVQLVECLSQRKWECLFCLNKCNCSACNRK
Dlac-CbfB       KATCHQCKLARPRNKIAYCHS-(36)-CQK---KYCSQCLWNRYQVNLVECLSVKKWECMFCRDKCNCSACNRK
Ppal-CbfB       KSTCHQCKLNKPLSKMAFCSG-(12)-CNK---KFCFHCLRNRYNLLVDCQKSLN-WLCLVCQGICNCGACVKK
Asub-CbfB       RSTCHQCKLPKPNSKMAQCTN-(03)-CGK---MYCGQCLATRYNTVLIDCQKTS-WLCLHCKGMCTCSTCLKK
Dfas-CbfB       KATCHQCKLARPRSRMAYCTR-(23)-CQK---KYCQQCLMNRYNMTLIDCLKEKKWECMFCRDRCDCSACIKR
Aspergillus     GPKCHICKYVEPEWKLASCSN-(02)-CSL---RYCYGSLFRAFNMLPQEVLETSKWECPRCQKFCSCGGCRKN
Neosartorya     GGRCHICKYIEPIWKLASCDY------CNQ---NYCYGSLFRAFNIQPHETMEVYRWMCPKCQKVCSCAACRRD
Coccidioides    HASCHICKTPEPKWKLATCDA------CGL---NYCYGSLYRAFDTSPREILEERSWRCPKCRKICSCGACRRD
Phaeosphaeria   HKSCHFCLHKHPKWKMAFCSS------CDL---SYCYGTLFRAHDMMPLSIMENRNWKCPHCRRVCNTGHCRRD
Neurospora      VYRCHVCQHKDYTYKLAFCSN-(02)-CSE---AYCYGVLYRAFDLMPQEVLQNERWECPRCLKICNCGACRRL
GIChaetonium    VHRCHVCQHKDYNYKLAFCTN-(02)-CHD---AYCFGVLYRAFDMMPQEVVQNERWQCPKCLKICNCAACRRA
Gibberella      LRRCHVCCHKDYAYRVHQCTN-(02)-CTE---SYCYGVLYRAFDMLPQKVLEDEQWQCPRCLGICNCGYCRRV
Magnaporthe     TYRCHVCCHRDNTFRLAFCSN-(02)-CAE---AYCYGTLYRAFDLMPQDVLRTENWKCPKCLKICNCGSCRKA
Naegleria       EVFCHQCQTHKSDYLFSECSK------CHLNYCNQCLWDRYALTLCGCKRVCN---WVCPKCNDACNCKNCLRS
Homo            GNTCHQCRQKTIDTKTV-CRN-(02)-CCGVRGQFCGPCLRNRYGEDVRSALLDPDWVCPPCRGICNCSYCRKR
Arabidopsis     GKTCHQCRQKTMGHRTQ-CSE------CNLVQGQFCGDCLFMRYGEHVLEALENPDWICPACRGICNCSLCRNN
```

Fig. S5: Comparison of DNA microarray and RNA-seq results.
Plotted are the expression levels of 12,317 genes in AX2 and JH.D cellsm measured in the RNA-seq experiment and expressed as RPKM values (**grey circles**). In this data set the 4,196 genes present on the DNA microarray (Lucas et al. PLoS ONE 2009) are labeled as **black circles**. Genes that were differentially regulated more than 1.5-fold in the microarray experiment are labeled with **red circles** and **green circles**, respectively.

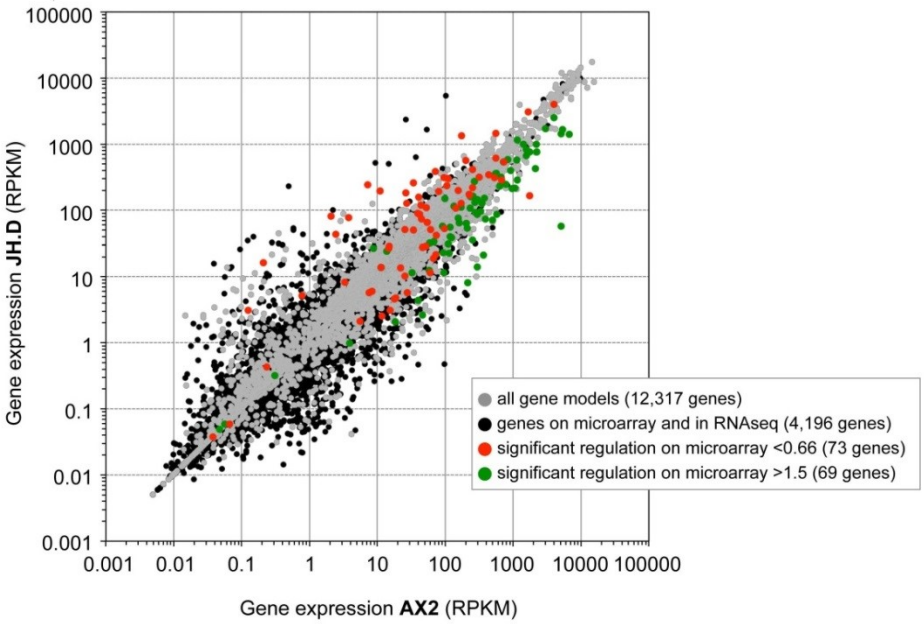


Table S1: List of primers used in this study.

Gene name	Primer name	Sequence (5'→3')
cyp513F1	DDB_G0278679-01	CACCATTATCTGAACCACATTATTGTTC
	DDB_G0278679-02	CTCTGCAAATGATTTACCAATACAACC
	DDB_G0270290-01	GCAATGTTTTCATTCGCAGTTTATGGTG
	DDB_G0270290-02	CATGTTGTAAATCTTTTGGAATTGCTTC
rdiB	DDB_G0280049-01	GGTTATATGGCAAGTGAACCAATCACAG
	DDB_G0280049-02	GTAAAGCATTTTCCCAATCTGGAGCAAC
	DDB_G0291033-01	CTAGTCAAAAAGACATACACTCAGAACC
	DDB_G0291033-02	CATCACTGAAATAGAATTACTGGTACC
agnC	DDB_G0271870-03	CTTGGTAATCCTGATCAAGGTGTTGTTG
	DDB_G0271870-04	GTGCACTTTTATGAGAGTATTGGCATAC
	DDB_G0270058-01	CATTCTACGTTCGTGTTATTGTAGATGG
	DDB_G0270058-02	CCATTTGCATCAACACCCCAAATTTGAC
abpF	DDB_G0291229-01	CAACTATAGTAATTTGGTTGATATTACC
	DDB_G0291229-02	CAGTGATATATACGGCAAGATTATTGG
	DDB_G0290885-01	GATTCAATGATAAAAGACCAAAGGGTGG
	DDB_G0290885-02	CATCAACAGTTGAATATACTCCAAGAG
	DDB_G0277333-01	CAAATGGTAGATTAGTTTTCATCGATAG
	DDB_G0277333-02	GTAGGTACTATTTATGATAACTGAATC
	DDB_G0291344-01	GGACTTTCTCATGAAATAAAGATATATG
	DDB_G0291344-02	CAAACTTGATCAGCACAATAATCATAAC
	DDB_G0289629-01	CTGCTGATGTGGAGGTCAATGGCAAAGG
	DDB_G0289629-02	AGTGATTGATAATTAATTGCTTTAACAC
kif12	DDB_G0268258-03	CTCACCATCTCCAACTAAACTTCAACC
	DDB_G0268258-04	AGAGGATTTACTAGAGGTGGAAGCTTC
wipA	DDB_G0270430-01	CCACCACCATTAATTGCCAAAAATACTG
	DDB_G0270430-02	GTGCAGGTGCTTGACGAGCACCAACTGG
ChLim	DDB_G0267490-03	GTTCACCATTCATTACTTCTGGATGGC
	DDB_G0267490-04	TTGAGCCCTCAATGATTGGTTTACTAC
ctxB	DDB_G0276893-01	GATGCCCTTCGTGGTCTCGATGATGAG
	DDB_G0276893-02	CCATTTATGAATATCAGAGAATTGAGC
myoB	DDB_G0289117-01	CAAGGTGGTGGTGCTCGTCCAATGCCAC
	DDB_G0289117-02	GGTTGTGGAGCTGGTTTAGCAACAGTTG
cpnA	DDB_G0293008-01	GCTTCTCAAGGCCAAAGCCAATCAAATC
	DDB_G0293008-02	CGATATCTCTTTCAGCTCTAACACCACC
proA	DDB_G0287125-01	GATGGTGGTGTTTGGGCTAAATCATCAG
	DDB_G0287125-02	CATCATAAATGCCAACAATGATTGCTTG
ctnnA	DDB_G0285939-03	GGCTATAATTGAAGAAGAAAAGAAACG
	DDB_G0285939-04	CATACCTTTAACATCACCATTCATTGC
vwkA	DDB_G0268144-01	GCAAAAGTCTTACAAATGAAACATTCTG
	DDB_G0268144-02	GTATCAGTGGAATGAATTGCTGGATCAG
aurK	DDB_G0279343-03	CTACTTATACATTACAAATTGCAGATG
	DDB_G0279343-04	CCAAACATCAGCTGTTTGATCATAACC
catA	DDB_G0274595-01	GTTTCGCTGCTCGTCAACCATACAATC
	DDB_G0274595-02	GCACGAACTTGAATTTCTTTGATGGTG

Table S2: List of CbfA-regulated genes.

Gene ID	Gene name	RPKM AX2	RPKM JH.D	RPKM JH.D[CbfA-CTD]	Fold change JH.D/AX2	Fold change JH.D[CbfA-CTD]/AX2	Gene description
DDB_G0283511		0.495	232.056	9.222	458.599	14.307	
DDB_G0278679	*cyp513F1*	0.039	15.620	0.434	388.170	8.555	cytochrome P450 family protein
DDB_G0270290		0.123	34.966	2.422	281.701	15.445	cellulose-binding domain-containing protein
DDB_G0280049	*rdiB*	0.095	26.015	0.874	280.247	7.469	rho GDP-dissociation inhibitor-like protein
DDB_G0291033		0.148	40.064	2.990	262.790	15.527	
DDB_G0271870	*agnC*	0.061	13.765	0.424	224.735	5.501	argonaut-like protein
DDB_G0293288	*psiQ*	0.071	14.074	0.126	201.160	1.429	PA14 domain-containing protein
DDB_G0270368		0.068	13.242	4.994	201.017	59.552	
DDB_G0280959	*DG1113*	0.015	2.604	0.077	174.696	4.058	C2 domain-containing protein
DDB_G0267716		0.056	9.855	2.541	170.974	34.870	
DDB_G0281691		0.352	45.056	1.141	128.091	2.574	
DDB_G0279727		0.279	32.372	1.770	115.058	4.990	
DDB_G0283595		0.019	1.935	0.172	98.148	6.883	patatin family protein
DDB_G0281391	*abcG5*	0.056	5.543	3.032	95.586	41.522	ABC transporter G family protein
DDB_G0268456		0.577	53.467	1.380	93.775	1.917	hssA/2C/7E family protein
DDB_G0289655	*abcG7*	0.027	2.632	0.121	92.186	3.313	ABC transporter G family protein
DDB_G0278069	*gacH*	0.027	2.425	0.161	88.864	4.677	RhoGAP domain-containing protein
DDB_G0267726	*uduA2*	26.246	2332.630	125.807	88.695	3.797	
DDB_G0277853	*ecmA*	0.039	3.438	0.410	87.436	8.292	extracellular matrix protein ST430
DDB_G0272975		0.104	9.363	1.104	86.429	8.044	
DDB_G0279935		0.576	50.432	8.279	85.772	11.145	C2 domain-containing protein
DDB_G0268272	*aslN*	0.076	6.409	0.347	84.362	3.615	AMP-dependent synthetase and ligase domain-containing protein
DDB_G0282643		0.349	31.428	1.106	83.980	2.337	
DDB_G0268398		0.208	16.391	1.254	77.792	4.700	hssA/2C/7E family protein
DDB_G0291674		0.083	6.068	1.489	75.792	14.723	short-chain dehydrogenase/reductase (SDR) family protein
DDB_G0284441		0.282	21.524	11.828	74.758	32.448	
DDB_G0291672		0.401	30.278	3.874	74.637	7.583	
DDB_G0267780		0.024	1.745	0.910	72.732	29.943	TRE3-C ORF2
DDB_G0279483	*pldB*	0.155	11.211	3.049	71.076	15.287	phospholipase D1
DDB_G0292732	*tgrF1*	0.030	1.842	0.093	59.362	2.384	IPT/TIG domain-containing protein
DDB_G0348179		9.298	520.563	50.789	56.793	4.387	
DDB_G0267398	*gppA*	0.446	25.683	4.863	56.278	8.458	glycoprotein gp100

DDB_G0280053	*osbA*	0.796	43.358	6.413	52.687	6.141	oxysterol binding family protein, member 1
DDB_G0267728	*uduA3*	102.142	5377.900	91.362	52.665	0.712	
DDB_G0279051		0.022	1.109	0.054	51.554	2.005	
DDB_G0284675		0.136	7.030	0.638	49.619	3.555	
DDB_G0285643	*mrkB*	0.059	2.944	0.241	49.285	3.197	MARK subfamily protein kinase
DDB_G0269266		0.018	0.893	0.350	48.362	15.055	TRE3-D ORF2
DDB_G0284233		0.233	11.084	1.352	48.285	4.666	
DDB_G0292792	*cyp508C1*	0.047	2.141	0.118	45.660	1.958	cytochrome P450 family protein
DDB_G0278875	*dduD*	2.207	100.234	6.714	45.371	2.405	
DDB_G0291902		0.222	9.630	2.192	43.331	7.743	cyclin-like F-box containing protein
DDB_G0274391	*alfA*	0.421	18.069	7.038	42.690	13.123	alpha-fucosidase
DDB_G0281035		0.254	10.424	2.611	39.504	7.774	
DDB_G0275117		0.214	8.568	3.703	39.349	13.454	
DDB_G0289925	*fmoG*	0.073	2.932	0.251	38.811	2.621	N,N-dimethylaniline,NADPH:oxygen oxidoreductase, N-oxide-forming
DDB_G0290837		0.372	14.930	0.709	38.577	1.447	
DDB_G0276203		0.397	15.488	1.272	37.791	2.453	
DDB_G0281387	*srfA*	0.138	5.389	1.522	37.726	8.439	SRF-related protein
DDB_G0286925		2.093	81.078	33.664	37.548	12.354	esterase/lipase/thioesterase domain-containing protein
DDB_G0290931		0.116	4.263	1.340	37.059	9.155	meprin and TRAF homology (MATH) domain-containing protein
DDB_G0289577		0.197	6.863	1.685	34.745	6.717	cyclin-like F-box containing protein
DDB_G0281083		0.078	2.681	0.320	34.704	3.287	
DDB_G0272706		0.098	3.466	0.594	34.336	4.606	terpenoid synthase domain-containing protein
DDB_G0279751		7.206	242.716	27.280	33.493	2.981	
DDB_G0283153	*cbpD1*	14.741	506.887	74.638	33.333	3.892	calcium-binding protein
DDB_G0278681		0.092	3.187	0.564	33.123	4.648	
DDB_G0292100	*fscH*	0.037	1.207	0.676	31.758	13.980	G-protein-coupled receptor (GPCR) family protein
DDB_G0293720		0.039	1.236	2.851	31.285	57.388	
DDB_G0270606		0.744	22.800	4.836	31.223	5.228	
DDB_G0268332	*uduA1*	53.331	1667.252	23.295	31.005	0.345	
DDB_G0277187		0.053	1.592	0.104	30.652	1.561	
DDB_G0294577		0.065	1.928	0.645	30.313	8.093	Rab GTPase
DDB_G0281513		0.030	0.871	0.072	29.687	1.917	equilibrative nucleoside transporter (ENT) family protein
DDB_G0283465		3.178	87.942	0.687	28.213	0.175	
DDB_G0284295	*iliG*	1.105	31.637	1.472	27.941	1.032	Endo-1,4-beta-glucanase family protein
DDB_G0270878		0.052	1.348	0.135	25.725	2.053	
DDB_G0286309		0.050	1.283	0.538	25.407	8.459	zinc-containing alcohol dehydrogenase (ADH)
DDB_G0274153		0.125	3.079	1.021	24.100	6.310	protein phosphatase 2C

DDB_G0282753		0.023	0.543	0.197	24.054	6.916	TRE3-D ORF2
DDB_G0282107		2.855	68.815	1.068	23.946	0.295	putative phospholipid transfer protein
DDB_G0284397		1.801	43.777	17.463	23.694	7.466	
DDB_G0269264		0.053	1.221	0.377	23.332	5.713	TRE3-C ORF2
DDB_G0289901		0.195	4.553	0.209	23.119	0.837	
DDB_G0285627		0.267	5.975	4.492	23.019	13.716	
DDB_G0281771		0.252	5.782	2.065	22.924	6.491	
DDB_G0284523		0.063	1.423	0.378	22.393	4.652	
DDB_G0287341		0.102	2.284	0.276	22.145	2.087	transmembrane protein
DDB_G0279521		0.093	2.080	1.678	21.763	13.878	
DDB_G0271586	*tgrK1*	0.016	0.336	0.110	21.640	5.596	immunoglobulin E-set domain-containing protein
DDB_G0283553		0.103	2.207	0.297	21.429	2.292	patatin family protein
DDB_G0268838		0.184	4.043	1.087	21.410	4.588	
DDB_G0274393		0.098	2.166	0.360	21.241	2.775	
DDB_G0267640	*gtaE*	0.904	18.908	5.450	21.137	4.805	GATA zinc finger domain-containing protein 5
DDB_G0291253	*dia2*	3.775	77.513	70.871	20.648	14.823	
DDB_G0292772	*tgrG1*	0.015	0.293	0.050	20.208	2.699	immunoglobulin E-set domain-containing protein
DDB_G0268400		0.211	4.250	1.673	19.959	6.149	hssA/2C/7E family protein
DDB_G0281853	*iliL*	0.344	6.685	2.280	18.977	5.131	putative cell surface glycoprotein
DDB_G0275575		0.052	0.995	0.078	18.899	1.196	
DDB_G0271438		0.064	1.238	2.262	18.873	27.194	
DDB_G0286055		0.155	2.881	0.563	18.859	2.917	peptidase C1A family protein
DDB_G0292890		0.163	2.964	0.356	18.697	1.753	
DDB_G0283887		0.134	2.389	0.267	18.451	1.631	
DDB_G0278197		1.271	23.199	1.469	17.887	0.892	
DDB_G0269206	*abcG21*	10.961	196.308	4.513	17.867	0.324	ABC transporter G family protein
DDB_G0289919	*ponJ*	0.111	1.998	1.649	17.846	11.641	ponticulin-related protein
DDB_G0293610		0.792	14.610	2.408	17.675	2.294	LMBR1-like conserved region-containing protein
DDB_G0293154		1.143	20.014	2.185	17.604	1.515	
DDB_G0268028		2.425	42.992	11.963	17.478	3.853	
DDB_G0278911		0.078	1.303	0.182	17.251	1.889	
DDB_G0292814	*symA*	36.766	633.415	5.779	17.021	0.123	solute carrier family 2 member protein
DDB_G0272714		0.070	1.150	0.140	16.986	1.626	
DDB_G0287971		0.082	1.356	0.382	16.634	3.716	
DDB_G0283345	*comD*	0.364	6.115	1.082	16.562	2.313	drug resistance transporter, EmrB/QacA subfamily protein
DDB_G0287695		1.822	28.925	1.674	16.542	0.761	FAD-binding monooxygenase
DDB_G0272386		0.680	10.970	5.697	16.534	6.789	cyclin-like F-box containing protein
DDB_G0267350		0.129	2.084	0.492	16.271	3.029	DDT-A
DDB_G0282635		0.128	1.963	0.468	15.858	2.985	

DDB_G0278489		0.064	0.986	0.273	15.651	3.390	
DDB_G0285771		0.047	0.771	0.208	15.569	3.322	
DDB_G0269940	*osbC*	0.053	0.789	0.211	15.216	3.248	oxysterol binding family protein, member 3
DDB_G0289625		0.802	12.015	5.277	15.215	5.267	
DDB_G0279719		0.015	0.227	0.069	15.111	3.682	putative protein serine/threonine kinase
DDB_G0289693		8.460	125.871	26.703	15.048	2.520	
DDB_G0279983	*ctrB*	0.404	6.324	1.560	14.982	2.905	solute carrier family 7 member protein
DDB_G0275327	*rabX*	0.080	1.233	0.353	14.973	3.410	Rab GTPase
DDB_G0283509		0.294	4.380	0.420	14.728	1.111	
DDB_G0275225		0.254	3.629	0.186	14.297	0.581	
DDB_G0280643		0.074	1.019	0.483	13.923	5.181	protein kinase, CMGC group
DDB_G0285781		0.444	5.951	0.970	13.279	1.707	
DDB_G0292844		0.556	7.577	2.261	13.132	3.097	
DDB_G0277141	*cotC*	0.213	2.753	0.568	13.116	2.148	spore coat protein SP60
DDB_G0293364	*sigB*	2.648	34.182	20.650	12.890	6.122	peptidase M8, leishmanolysin family protein
DDB_G0271160		0.039	0.506	0.854	12.723	17.056	
DDB_G0270292		0.116	1.441	0.194	12.626	1.351	cellulose-binding domain-containing protein
DDB_G0293578		0.022	0.270	0.135	12.582	4.974	
DDB_G0281321		0.310	3.788	0.307	12.240	0.780	
DDB_G0288185		0.157	1.891	0.325	12.218	1.656	hssA/2C/7E family protein
DDB_G0284729	*fslK*	0.957	11.252	1.474	12.071	1.254	G-protein-coupled receptor (GPCR) family protein
DDB_G0278719		0.039	0.452	0.054	11.955	1.125	immunoglobulin E-set domain-containing protein
DDB_G0290191		1.386	16.228	6.377	11.817	3.671	
DDB_G0268172		0.164	1.865	0.384	11.532	1.867	hssA/2C/7E family protein
DDB_G0276297		0.426	5.023	1.471	11.526	2.674	
DDB_G0283549		3.490	40.505	30.168	11.382	6.706	C2 domain-containing protein
DDB_G0270582		0.648	7.174	2.096	11.148	2.568	putative carboxypeptidase
DDB_G0280943	*elmoD*	0.205	2.295	0.397	10.983	1.494	ankyrin repeat-containing protein
DDB_G0280957		0.157	1.700	0.235	10.913	1.200	
DDB_G0275063		0.109	1.212	0.133	10.801	0.934	
DDB_G0293992		3.398	35.271	42.244	10.772	10.235	C2 domain-containing protein
DDB_G0284465	*cudA*	1.601	16.850	10.127	10.713	5.088	putative transcriptional regulator
DDB_G0272014		1.239	13.270	39.488	10.696	25.167	pyridoxal-phosphate-dependent aminotransferase
DDB_G0282715		0.423	4.523	7.132	10.681	13.265	Neutral and basic amino acid transport protein rBAT
DDB_G0275171		0.744	7.883	1.633	10.633	1.740	
DDB_G0269008		0.187	2.023	1.054	10.469	4.348	
DDB_G0276285		1.273	13.492	21.567	10.430	13.163	
DDB_G0282283	*cyp515B1*	0.505	5.147	1.300	10.232	2.043	cytochrome P450 family protein
DDB_G0276261	*mcfV*	6.514	67.513	15.957	10.116	1.898	mitochondrial substrate carrier family protein

DDB_G0269618		0.173	1.723	0.623	10.085	2.887	
DDB_G0277147	*stkA*	1.432	14.122	13.361	9.966	7.444	GATA zinc finger domain-containing protein 1
DDB_G0289121	*pde4*	0.190	1.899	0.679	9.897	2.800	cAMP-specific phosphodiesterase
DDB_G0267684		0.043	0.442	0.138	9.809	2.433	putative cell surface glycoprotein
DDB_G0290633		0.049	0.474	0.168	9.789	2.732	
DDB_G0283287	*fut7*	0.091	0.876	0.382	9.744	3.332	alpha-3/4-fucosyltransferase
DDB_G0293334		0.041	0.388	1.053	9.732	21.028	
DDB_G0287955		0.097	0.939	0.179	9.658	1.453	
DDB_G0283305		1.394	13.216	3.376	9.324	1.882	ERV/ALR sulphydryl oxidase domain-containing protein
DDB_G0270750		4.921	44.860	13.084	9.173	2.120	Kelch repeat-containing protein
DDB_G0286897		4.600	41.370	1.427	9.154	0.250	
DDB_G0286413		0.210	1.814	0.595	8.809	2.296	
DDB_G0290469	*pks28*	0.255	2.290	0.716	8.709	2.160	beta-ketoacyl synthase family protein
DDB_G0284759	*psiM*	0.450	3.955	2.764	8.632	4.781	PA14 domain-containing protein
DDB_G0275223		0.193	1.576	0.190	8.412	0.802	
DDB_G0295791		1.807	15.721	3.968	8.371	1.673	
DDB_G0270926		0.299	2.555	0.120	8.291	0.312	
DDB_G0269268		0.537	4.586	2.320	8.287	3.308	TRE3-D ORF1 fragment
DDB_G0274385		0.125	1.028	0.905	8.229	5.713	Cysteine proteinase 1, mitochondrial
DDB_G0270198	*hcpC*	0.246	2.048	0.768	8.209	2.434	HP1-like protein
DDB_G0286873		0.113	0.923	0.096	8.187	0.674	
DDB_G0281379		0.377	3.153	3.569	8.161	7.268	
DDB_G0278953		2.670	21.362	8.178	8.084	2.452	CBS (cystathionine-beta-synthase) domain-containing protein
DDB_G0283899	*mgp4*	5.591	44.391	23.147	8.036	3.317	Cdc15/Fes/CIP4 domain-containing protein
DDB_G0290815		0.753	6.239	0.568	7.985	0.577	
DDB_G0290161	*miox*	1.199	9.592	11.875	7.904	7.750	myo-inositol oxygenase
DDB_G0287809		0.182	1.440	0.278	7.839	1.208	
DDB_G0293182	*expl8*	0.373	2.907	0.453	7.643	0.941	expansin-like protein
DDB_G0292546		0.218	1.682	0.832	7.612	2.960	
DDB_G0285959		0.101	0.736	0.154	7.568	1.244	
DDB_G0267528		173.869	1332.233	221.780	7.564	0.998	
DDB_G0270218	*glkA*	0.173	1.326	0.817	7.443	3.618	GSK family protein kinase
DDB_G0286853		34.156	259.254	72.026	7.405	1.628	
DDB_G0290543		0.392	2.829	1.309	7.402	2.695	
DDB_G0272674		0.017	0.128	0.062	7.268	2.771	
DDB_G0275145		0.430	3.180	2.068	7.249	3.729	IBR zinc finger-containing protein
DDB_G0283675		0.398	2.830	0.805	7.215	1.615	
DDB_G0269450		1.760	12.578	11.573	7.163	5.207	

DDB_G0284743	*4cl3*	0.235	1.644	0.619	7.031	2.098	4-coumarate-CoA ligase
DDB_G0269208	*abcG19*	0.427	3.080	0.784	7.004	1.410	ABC transporter G family protein
DDB_G0277265		7.066	49.415	10.135	6.977	1.129	
DDB_G0274941	*omt2*	0.085	0.608	0.232	6.933	2.111	O-methyltransferase family 2 protein
DDB_G0279897		0.537	3.622	4.423	6.902	6.704	
DDB_G0293788		0.770	5.164	1.962	6.877	2.054	
DDB_G0268696		26.068	182.340	53.029	6.865	1.575	
DDB_G0267324		0.106	0.743	0.638	6.778	4.617	DIRS1 ORF1
DDB_G0293886	*omt11*	0.099	0.638	0.073	6.747	0.610	O-methyltransferase family 2 protein
DDB_G0280817		0.506	3.574	0.751	6.745	1.123	
DDB_G0269584		1.945	13.177	2.162	6.645	0.861	
DDB_G0283431		1.487	10.217	6.159	6.546	3.124	
DDB_G0284819		0.550	3.760	0.565	6.511	0.773	
DDB_G0276235		0.148	0.976	0.895	6.499	4.713	
DDB_G0286119	*tagB*	0.815	5.405	2.695	6.493	2.552	ABC transporter B family protein
DDB_G0278531		2.407	15.398	2.230	6.492	0.743	
DDB_G0292822		0.051	0.330	0.216	6.480	3.385	
DDB_G0290695		0.285	1.912	0.633	6.438	1.684	
DDB_G0291946		0.213	1.322	0.112	6.429	0.430	
DDB_G0286243		1.439	9.293	1.425	6.375	0.773	
DDB_G0271598	*tgrS1*	0.074	0.454	0.086	6.374	0.956	immunoglobulin E-set domain-containing protein
DDB_G0272218		1.709	10.916	4.094	6.346	1.879	
DDB_G0282137		0.598	3.682	0.805	6.265	1.088	
DDB_G0291786		0.279	1.712	0.452	6.187	1.292	
DDB_G0277105		0.070	0.449	0.131	6.174	1.429	
DDB_G0288967		0.158	1.027	0.905	6.165	4.333	
DDB_G0270000		0.407	2.558	2.632	6.072	4.919	
DDB_G0278363		0.619	3.798	2.378	6.035	2.995	
DDB_G0287583	*gerC*	2.085	12.022	10.278	5.885	3.970	germination protein p109
DDB_G0281149		16.243	96.698	13.754	5.878	0.659	
DDB_G0276087		1.423	8.347	7.256	5.820	3.980	Monoglyceride lipase
DDB_G0282111		0.107	0.622	0.183	5.759	1.351	
DDB_G0286643	*grlM*	2.069	12.384	5.018	5.750	1.843	G-protein-coupled receptor (GPCR) family 3 protein 12
DDB_G0293738	*cyp521A1*	1.594	8.856	2.123	5.653	1.066	cytochrome P450 family protein
DDB_G0280973	*abcC12*	0.388	2.286	0.530	5.583	1.018	ABC transporter C family protein
DDB_G0287963		8.515	46.967	1.000	5.554	0.093	
DDB_G0282491		6.423	36.136	6.864	5.503	0.825	WD-40 repeat-containing protein
DDB_G0269914	*tpp1*	1.255	7.218	2.752	5.461	1.644	peptidase S8 and S53 domain-containing protein
DDB_G0284111		59.825	330.858	65.855	5.385	0.845	

DDB_G0270328		0.261	1.397	0.941	5.367	2.850	
DDB_G0280533	*lmcB*	71.048	383.769	1509.206	5.349	16.668	
DDB_G0290097		0.188	1.040	0.441	5.348	1.784	
DDB_G0270888		1.005	5.380	2.242	5.332	1.763	TRE3-D ORF1
DDB_G0286321		0.347	1.859	0.771	5.321	1.753	putative glycoside hydrolase
DDB_G0282629		1.010	5.406	2.681	5.313	2.088	
DDB_G0287705		0.749	3.950	0.704	5.312	0.746	
DDB_G0289581		1.608	8.524	5.799	5.292	2.846	GCN5-related N-acetyltransferase
DDB_G0268828		72.291	384.847	170.650	5.283	1.857	
DDB_G0270588		0.586	3.081	5.119	5.281	6.917	
DDB_G0270436		7.276	39.066	20.508	5.270	2.179	
DDB_G0272376		0.225	1.193	0.834	5.233	2.910	
DDB_G0291622	*tgrA1*	0.308	1.588	0.118	5.182	0.303	immunoglobulin E-set domain-containing protein
DDB_G0267624		0.933	4.849	3.433	5.171	2.915	
DDB_G0268854	*gnt4*	3.103	16.227	5.746	5.144	1.438	alpha-1,3-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase
DDB_G0267562		0.773	3.957	2.674	5.111	2.729	
DDB_G0292246		9.248	47.670	8.362	5.100	0.706	
DDB_G0278127		49.281	254.260	4.430	5.082	0.070	
DDB_G0276125		0.247	1.289	0.265	5.071	0.829	
DDB_G0293286	*psiR*	4.813	25.034	5.707	5.060	0.912	PA14 domain-containing protein
DDB_G0295769		20.781	102.872	11.566	4.979	0.439	
DDB_G0293978	*gxcJ*	0.465	2.283	1.916	4.936	3.265	RhoGEF domain-containing protein
DDB_G0272150	*grlJ*	2.720	14.289	8.474	4.916	2.299	G-protein-coupled receptor (GPCR) family 3 protein 9
DDB_G0290317	*psiJ*	0.278	1.364	0.634	4.879	1.789	PA14 domain-containing protein
DDB_G0268812		0.166	0.867	0.280	4.869	1.240	putative AFK family protein kinase
DDB_G0268928		56.118	274.836	37.389	4.868	0.523	phosphoesterase, PA-phosphatase related-family protein
DDB_G0288813		0.054	0.260	0.231	4.858	3.397	EGF-like domain-containing protein
DDB_G0268012		1.801	8.858	3.008	4.852	1.302	response regulator receiver domain-containing protein
DDB_G0292412		7.388	36.749	8.664	4.847	0.901	putative potassium transporter
DDB_G0278405		0.213	1.031	0.798	4.841	2.950	
DDB_G0283419	*gpaI*	26.905	127.856	34.524	4.797	1.023	G-protein subunit alpha 9
DDB_G0271640		1.704	8.420	1.888	4.796	0.850	
DDB_G0291354		8.403	41.960	3.540	4.780	0.318	
DDB_G0287299		0.587	2.853	0.723	4.725	0.946	
DDB_G0291117	*noxC*	0.992	4.682	0.641	4.701	0.506	superoxide-generating NADPH oxidase flavocytochrome
DDB_G0288645		13.082	62.754	23.064	4.696	1.359	
DDB_G0289073	*csaA*	2.689	12.604	8.594	4.693	2.542	contact site A protein
DDB_G0287673		4.090	19.516	3.299	4.688	0.623	

DDB_G0283317	*tgrI2*	0.206	0.984	0.308	4.641	1.144	immunoglobulin E-set domain-containing protein
DDB_G0293840	*drcA*	2.396	11.369	2.090	4.634	0.673	putative glutathione transferase
DDB_G0291938		1.116	5.221	2.094	4.633	1.464	
DDB_G0286411	*bzpN*	4.702	22.007	9.989	4.613	1.656	putative basic-leucine zipper (bZIP) transcription factor
DDB_G0282407		2.615	12.333	16.544	4.609	4.899	
DDB_G0267206		1.160	5.618	3.378	4.606	2.193	DIRS1 ORF2 fragment
DDB_G0288195	*sibC*	29.797	136.409	78.820	4.566	2.085	type A von Willebrand factor (VWFA) domain-containing protein
DDB_G0279859		0.099	0.468	1.904	4.563	14.749	EGF-like domain-containing protein
DDB_G0295741		3.157	14.701	2.967	4.551	0.722	putative acyl-CoA N-acyltransferase
DDB_G0280337		0.778	3.683	0.868	4.534	0.850	
DDB_G0278655		2.024	9.008	2.743	4.509	1.075	short-chain dehydrogenase/reductase (SDR) family protein
DDB_G0288197	*sibD*	0.976	4.532	2.401	4.505	1.888	
DDB_G0275951		0.520	2.406	0.561	4.504	0.832	putative transmembrane protein
DDB_G0276245		1.369	6.281	0.881	4.503	0.499	acyl-CoA oxidase
DDB_G0285609		2.235	10.096	3.607	4.501	1.271	
DDB_G0288147		1.171	5.306	2.396	4.479	1.596	PKC domain-containing protein, PE/DAG binding
DDB_G0291952		3.956	17.850	40.217	4.471	7.891	
DDB_G0291666	*tgrA2*	0.130	0.593	0.111	4.467	0.659	immunoglobulin E-set domain-containing protein
DDB_G0269364	*stlA*	6.131	27.014	4.847	4.465	0.634	beta-ketoacyl synthase family protein
DDB_G0267182		0.515	2.409	1.582	4.461	2.311	DIRS1 ORF2 fragment
DDB_G0277023		2.206	9.924	0.510	4.460	0.180	
DDB_G0272214		0.128	0.570	1.112	4.425	6.833	
DDB_G0280741		1.587	7.097	1.581	4.419	0.776	
DDB_G0294286		0.625	2.863	1.383	4.360	1.668	DIRS1 ORF2 fragment
DDB_G0289861		3.365	14.796	2.662	4.358	0.619	
DDB_G0270886		0.292	1.280	0.757	4.333	2.028	TRE3-D ORF2
DDB_G0270540		2.099	9.031	2.348	4.325	0.887	
DDB_G0278181		8.296	36.374	15.160	4.325	1.430	putative GTPase activating protein (GAP)
DDB_G0274149		0.458	1.986	1.238	4.322	2.127	
DDB_G0274369		2.502	10.720	10.958	4.307	3.491	
DDB_G0267474	*sigD*	0.600	2.602	0.549	4.277	0.717	spore coat protein
DDB_G0279943		22.026	99.049	48.066	4.277	1.639	putative sodium-dependent phosphate transporter
DDB_G0286121	*tagC*	1.926	8.380	2.655	4.266	1.069	ABC transporter B family protein
DDB_G0284591		1.658	7.115	3.642	4.263	1.724	
DDB_G0267300		0.258	1.121	0.729	4.220	2.164	DIRS1 ORF1/ORF2 fusion fragment
DDB_G0286657		84.878	362.149	75.502	4.216	0.696	
DDB_G0290475		16.664	71.387	15.948	4.199	0.742	
DDB_G0272364		2.470	10.055	3.161	4.195	1.041	EGF-like domain-containing protein

DDB_G0267188		1.017	4.254	2.429	4.190	1.881	DIRS1 ORF2 fragment
DDB_G0278787		0.688	2.895	3.317	4.150	3.740	
DDB_G0274831		0.527	2.114	2.525	4.130	3.880	putative T4-like lysozyme
DDB_G0267666	*gefW*	4.907	20.946	8.926	4.124	1.390	RasGEF domain-containing protein
DDB_G0274117	*abcG8*	0.303	1.257	1.903	4.092	4.918	ABC transporter G family protein
DDB_G0276875	*hspG5*	0.617	2.462	31.317	4.045	40.648	heat shock protein Hsp20 domain-containing protein
DDB_G0307694		0.497	2.035	0.927	4.044	1.454	
DDB_G0280837		3.983	16.070	2.773	4.039	0.549	
DDB_G0276887	*fbxA*	10.556	42.498	16.217	4.017	1.212	START domain-containing protein
DDB_G0277045		1.681	6.728	3.305	4.006	1.566	
DDB_G0275669		23.868	100.275	10.549	4.003	0.332	proline dehydrogenase
DDB_G0267184		2.804	11.495	6.629	4.002	1.823	DIRS1 ORF2 fragment
DDB_G0267262		0.742	3.029	1.761	3.997	1.826	DIRS1 ORF2 fragment
DDB_G0276949	*hspG6*	2.266	8.964	115.178	3.991	40.594	heat shock protein Hsp20 domain-containing protein
DDB_G0274257	*fhit*	16.680	66.090	64.348	3.988	3.081	fragile histidine triad protein
DDB_G0276533		0.467	1.853	0.557	3.973	0.930	
DDB_G0294344		0.414	1.717	0.951	3.971	1.733	DIRS1 ORF2 fragment
DDB_G0293344		0.941	3.743	2.985	3.950	2.487	TRE5-C ORF2 fragment
DDB_G0283571		4.795	18.412	13.368	3.941	2.267	
DDB_G0267578		0.545	2.146	2.036	3.922	2.947	
DDB_G0274821		0.299	1.169	0.648	3.889	1.693	kinase motif-containing (KMC) protein
DDB_G0267210		0.592	2.370	1.460	3.883	1.887	DIRS1 ORF2 fragment
DDB_G0305708		0.973	3.968	2.007	3.882	1.555	DIRS1 ORF1 fragment
DDB_G0281653		24.392	97.754	24.640	3.875	0.770	
DDB_G0280479	*arrE*	23.708	95.185	39.431	3.873	1.268	ADP-ribosylation factor-related
DDB_G0279799	*cprB*	1.990	7.681	9.872	3.873	3.918	cysteine proteinase 2
DDB_G0284745	*4cl2*	1.641	6.493	3.045	3.869	1.432	4-coumarate-CoA ligase
DDB_G0287515	*tat*	40.756	158.807	19.992	3.863	0.385	tyrosine transaminase
DDB_G0269126		9.008	35.093	3.504	3.825	0.302	
DDB_G0281619		0.290	1.159	0.814	3.811	2.098	tRNA pseudouridine synthase
DDB_G0279121		1.789	6.934	4.960	3.802	2.171	
DDB_G0284109		108.171	411.976	43.935	3.798	0.321	
DDB_G0291217		0.908	3.560	2.079	3.793	1.744	DIRS1 ORF2 fragment
DDB_G0289871		1.185	4.591	1.648	3.786	1.074	
DDB_G0287863		0.071	0.261	0.173	3.780	1.977	
DDB_G0267244		0.136	0.536	0.446	3.747	2.456	DIRS1 ORF2 fragment
DDB_G0271996		2.141	8.042	5.480	3.740	2.013	FNIP repeat-containing protein
DDB_G0291105	*cyp513C1*	15.958	57.905	6.490	3.704	0.330	cytochrome P450 family protein
DDB_G0268734		1.520	5.661	0.409	3.685	0.211	

DDB_G0287427		1.031	3.786	1.153	3.675	0.884	
DDB_G0271354		0.733	2.656	1.347	3.675	1.479	
DDB_G0294346		1.716	6.431	3.320	3.674	1.506	DIRS1 ORF1/ORF3 fusion fragment
DDB_G0288017		19.185	71.329	24.948	3.674	1.012	putative sphingomyelinase
DDB_G0272434		10.826	41.541	42.357	3.673	2.957	
DDB_G0287031	*gpaC*	7.451	28.700	13.256	3.656	1.334	G-protein subunit alpha 3
DDB_G0290655		17.050	61.498	28.067	3.651	1.313	
DDB_G0293802		2.176	8.107	2.998	3.641	1.064	transmembrane protein
DDB_G0267328		0.405	1.483	1.070	3.634	2.069	DIRS1 ORF2 fragment
DDB_G0285123		0.701	2.584	1.612	3.632	1.790	
DDB_G0291221		0.523	2.011	1.325	3.628	1.898	DIRS1 ORF1
DDB_G0291065		0.531	1.961	0.639	3.627	0.936	
DDB_G0268640		0.266	0.951	1.427	3.621	4.284	
DDB_G0283113	*eriA*	2.273	8.380	23.132	3.615	7.864	putative RNase III
DDB_G0267326		0.372	1.389	0.952	3.614	1.960	DIRS1 ORF2 fragment
DDB_G0286851	*tgrE1*	3.009	11.049	2.634	3.607	0.680	IPT/TIG domain-containing protein
DDB_G0292094		4.460	15.855	16.015	3.573	2.860	
DDB_G0280573		0.532	1.905	0.896	3.564	1.333	
DDB_G0275173	*hbx2*	2.269	8.148	14.667	3.550	5.074	putative homeobox transcription factor
DDB_G0270232		5.245	19.562	6.726	3.549	0.960	
DDB_G0267296		1.040	3.791	2.769	3.541	2.043	DIRS1 ORF2 fragment
DDB_G0288405		27.148	97.280	33.490	3.539	0.963	putative GTPase activating protein (GAP)
DDB_G0287511		2.062	7.362	3.981	3.519	1.500	
DDB_G0269040		18.442	65.881	66.659	3.519	2.821	IPT/TIG domain-containing protein
DDB_G0278487		137.684	496.401	143.063	3.510	0.799	CDK family protein kinase
DDB_G0274199		3.756	13.217	17.419	3.507	3.641	putative metallophosphoesterase
DDB_G0305706		0.361	1.308	0.853	3.489	1.789	DIRS1 ORF1 fragment
DDB_G0281507	*colC*	0.142	0.494	0.506	3.477	2.835	colossin C
DDB_G0276065		145.777	507.211	73.287	3.466	0.397	acyl-CoA oxidase
DDB_G0267282		0.993	3.517	2.538	3.462	1.964	DIRS1 ORF1 fragment
DDB_G0267272		4.951	16.831	9.266	3.439	1.489	DIRS1 ORF2 fragment
DDB_G0268454		4.845	16.211	8.766	3.431	1.460	
DDB_G0286141		58.846	200.687	145.430	3.426	1.966	
DDB_G0276119	*cnrF*	21.013	72.805	16.382	3.393	0.606	recoverin family protein
DDB_G0285219		21.544	73.611	29.953	3.391	1.091	DUF3430 family protein
DDB_G0285611		4.659	16.247	7.003	3.390	1.155	
DDB_G0269446		1.110	3.752	3.180	3.385	2.264	
DDB_G0285675		0.849	2.863	1.324	3.384	1.228	
DDB_G0274481		99.280	345.096	100.802	3.377	0.779	EXS domain-containing protein

DDB_G0291936		0.975	3.482	1.036	3.376	0.793	
DDB_G0270376		1.742	5.856	3.429	3.367	1.559	hssA/2C/7E family protein
DDB_G0267258		0.849	2.945	1.998	3.358	1.797	DIRS1 ORF2 fragment
DDB_G0279533		0.233	0.749	0.688	3.340	2.423	
DDB_G0283451		12.452	41.396	21.319	3.339	1.360	Phosphatase DCR2
DDB_G0287111		1.640	5.788	1.026	3.336	0.467	
DDB_G0294180		1.248	4.310	2.641	3.336	1.615	DIRS1 ORF1/ORF3 fusion fragment
DDB_G0294216		2.780	9.658	6.053	3.332	1.649	DIRS1 ORF2 fragment
DDB_G0267314		3.200	10.904	6.416	3.330	1.550	DIRS1 ORF2/ORF3 fusion fragment
DDB_G0267240		0.415	1.423	0.849	3.327	1.567	DIRS1 ORF2 fragment
DDB_G0280829		0.362	1.227	0.607	3.320	1.293	
DDB_G0294280		0.362	1.197	0.684	3.319	1.504	DIRS1 ORF2 fragment
DDB_G0288459		5.586	18.787	6.561	3.317	0.919	
DDB_G0286467		2.191	7.555	6.634	3.305	2.296	
DDB_G0280339		2.466	8.148	3.574	3.300	1.142	
DDB_G0283485		9.243	30.338	10.339	3.293	0.886	
DDB_G0269882		0.796	2.632	0.650	3.293	0.646	lipocalin family protein
DDB_G0290907		2.290	7.485	5.034	3.286	1.745	
DDB_G0282501		0.454	1.482	1.095	3.276	1.915	
DDB_G0280145		50.973	172.362	45.649	3.273	0.683	glycoside hydrolase family 63 protein
DDB_G0269574	*dcd3B*	80.234	270.191	65.557	3.271	0.626	N-acylsphingosine amidohydrolase
DDB_G0274661		188.454	653.931	56.734	3.266	0.224	
DDB_G0275855		0.652	2.137	1.136	3.256	1.370	
DDB_G0288331	*expl7*	0.505	1.663	2.222	3.237	3.414	expansin-like protein
DDB_G0284619		2.645	8.578	2.479	3.235	0.743	
DDB_G0277525		3.959	12.785	4.665	3.231	0.934	
DDB_G0275343		6.560	21.384	7.452	3.230	0.889	
DDB_G0294210		1.165	3.836	2.118	3.221	1.401	DIRS1 ORF2 fragment
DDB_G0291223		15.005	49.572	30.915	3.219	1.590	DIRS1 ORF2/ORF3 fusion fragment
DDB_G0287217		0.563	1.823	0.838	3.218	1.172	
DDB_G0270166		5.225	17.052	11.284	3.199	1.675	
DDB_G0292756		1.639	5.306	3.749	3.196	1.788	
DDB_G0275689	*abcG2*	95.916	311.297	78.993	3.182	0.638	ABC transporter G family protein
DDB_G0272030		5.475	17.382	220.149	3.180	31.819	
DDB_G0288825		3.044	10.005	3.929	3.171	0.980	
DDB_G0276965	*gltA*	44.178	143.518	47.470	3.162	0.828	citrate synthase
DDB_G0287587	*smlA*	1243.046	3902.803	1886.954	3.157	1.208	
DDB_G0282701		9.542	30.320	58.776	3.138	4.837	
DDB_G0274277		0.166	0.518	0.189	3.136	0.911	

DDB_G0290005		3.286	10.444	5.349	3.126	1.267	myotubularin-related protein
DDB_G0283807		8.807	27.656	8.136	3.124	0.725	
DDB_G0285911		8.102	25.741	13.904	3.121	1.333	LIM-type zinc finger-containing protein
DDB_G0288463	*gnt3*	1.550	4.899	1.146	3.120	0.574	alpha-1,3-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase
DDB_G0279913	*dhkH*	3.302	10.463	5.416	3.119	1.272	HisK family protein kinase
DDB_G0280997		1.567	5.040	4.567	3.109	2.226	zinc-containing alcohol dehydrogenase (ADH)
DDB_G0278651	*manE*	3.661	11.416	8.701	3.106	1.875	alpha-mannosidase
DDB_G0274973		0.296	0.912	0.625	3.089	1.666	
DDB_G0270296	*dcd1A*	50.104	160.809	68.018	3.086	1.034	acid N-acylsphingosine amidohydrolase-like protein
DDB_G0275679	*gxcJJ*	36.706	113.796	48.448	3.075	1.036	RhoGEF domain-containing protein
DDB_G0272158	*wrnip1*	1.656	5.183	6.255	3.058	2.920	AAA ATPase domain-containing protein
DDB_G0292560	*racJ*	6.860	21.277	9.632	3.055	1.091	Rho GTPase
DDB_G0267236		33.823	104.688	59.479	3.054	1.372	DIRS1 ORF2 fragment
DDB_G0268994	*manC*	0.883	2.684	1.184	3.049	1.053	alpha-mannosidase
DDB_G0284365		8.696	26.677	7.115	3.036	0.640	
DDB_G0271636		2.610	7.905	3.044	3.035	0.922	
DDB_G0279133	*syn7B*	72.288	223.742	96.790	3.035	1.036	putative syntaxin 7
DDB_G0277695		0.777	2.368	1.810	3.031	1.838	Uncharacterized protein yegU
DDB_G0267238		16.632	51.828	30.338	3.030	1.403	DIRS1 ORF1
DDB_G0280705		12.464	38.485	23.779	3.028	1.481	SNF2-related domain-containing protein
DDB_G0292696	*colA*	0.355	1.135	1.389	3.027	2.932	colossin A
DDB_G0275959		34.022	106.175	62.687	3.026	1.414	cellular retinaldehyde-binding/triple function domain-containing protein
DDB_G0280725		14.111	43.638	10.150	3.021	0.558	pirin family protein
DDB_G0276935		6.153	18.716	7.052	3.017	0.898	DDHD domain-containing protein
DDB_G0283149	*cotD*	3.147	9.333	3.820	3.016	0.978	spore coat protein SP75
DDB_G0272280		11.253	34.005	50.588	3.015	3.531	AhpC/TSA family protein
DDB_G0270934		10.509	31.582	15.844	3.006	1.194	
DDB_G0286067		1.640	4.968	1.398	3.001	0.669	
DDB_G0288307	*cnrM*	4.851	15.208	5.757	3.000	0.899	putative cell number regulator
DDB_G0290593		1577.297	527.734	1482.003	0.333	0.743	
DDB_G0294066	*mrpl14*	467.785	160.135	652.864	0.331	1.070	ribosomal protein L14, mitochondrial
DDB_G0289923	*agxt*	168.439	55.622	231.774	0.330	1.091	alanine-glyoxylate aminotransferase
DDB_G0293206		4.389	1.436	2.575	0.330	0.464	
DDB_G0277175	*forG*	14.760	4.921	10.794	0.329	0.571	formin homology domain-containing protein
DDB_G0294062	*mrps14*	336.433	114.328	497.828	0.329	1.136	ribosomal protein S14, mitochondrial
DDB_G0286247	*ponB*	4.653	1.535	115.229	0.328	19.565	ponticulin-related protein
DDB_G0281703	*nedd8l1*	4.405	1.496	6.150	0.328	1.081	neddylin-like protein

DDB_G0270340	*rapC*	43.194	14.482	75.528	0.328	1.354	Ras GTPase
DDB_G0294076	*DidioMp21*	73.843	24.536	94.415	0.325	0.993	
DDB_G0278865		8.457	2.765	41.135	0.325	3.853	
DDB_G0279785	*cniB*	10.622	3.429	17.960	0.324	1.338	cornichon family protein
DDB_G0282595	*gmfA*	67.226	22.371	95.291	0.324	1.089	ADF/cofilin-like domain-containing protein
DDB_G0279953		6.083	2.006	85.233	0.324	10.942	
DDB_G0279459		27.019	8.915	67.734	0.322	1.940	
DDB_G0271374		63.386	20.539	37.505	0.322	0.467	Glycerophosphodiester phosphodiesterase domain-containing protein 1
DDB_G0287893		5.578	1.808	7.961	0.322	1.122	
DDB_G0281139		11.179	3.627	31.768	0.321	2.226	
DDB_G0281371		4.212	1.366	4.601	0.321	0.860	
DDB_G0283773		11.799	3.861	17.727	0.320	1.166	putative GTP-binding protein
DDB_G0272174		9.665	3.152	11.458	0.320	0.917	
DDB_G0274169		8.466	2.688	11.040	0.319	1.034	bolA family protein
DDB_G0276859		1.511	0.484	2.809	0.319	1.466	
DDB_G0277759	*phyA*	3.536	1.125	4.987	0.319	1.109	prolyl 4-hydroxylase
DDB_G0289969		21.015	6.709	1.873	0.318	0.071	
DDB_G0278029		1.425	0.451	1.872	0.318	1.039	
DDB_G0287097		29.499	9.504	59.226	0.317	1.569	FNIP repeat-containing protein
DDB_G0284637		0.731	0.235	1.389	0.317	1.489	
DDB_G0269790	*gaa*	284.990	89.960	293.095	0.317	0.819	alpha-glucoside hydrolase
DDB_G0290833	*rabK1*	1.956	0.630	3.154	0.317	1.254	Rab GTPase
DDB_G0286163		3.771	1.188	26.734	0.317	5.669	B-box zinc finger-containing protein
DDB_G0275981		46.887	15.343	116.538	0.317	1.911	
DDB_G0275731		10.516	3.344	18.545	0.317	1.394	calponin homology (CH) domain-containing protein
DDB_G0267414	*repD*	18.679	6.004	26.274	0.316	1.100	transcription factor IIH component
DDB_G0293726		13.653	4.442	35.648	0.315	1.986	TRE5-A ORF1
DDB_G0288873	*pich*	12.692	4.151	28.589	0.315	1.720	polo-interacting checkpoint helicase
DDB_G0277701	*apeA*	236.632	76.877	85.878	0.314	0.279	DNA-(apurinic or apyrimidinic site) lyase
DDB_G0269376		7.675	2.429	6.346	0.314	0.647	
DDB_G0277699		33.926	10.696	11.414	0.313	0.264	
DDB_G0277765		97.452	31.097	48.839	0.312	0.389	Skipper GAG-PRO
DDB_G0291402		129.201	40.978	128.125	0.312	0.774	Osteoclast-stimulating factor 1
DDB_G0293714		124.326	38.561	86.210	0.311	0.550	
DDB_G0290111	*adhfe1*	21.797	6.751	27.887	0.311	1.022	alcohol dehydrogenase iron-containing 1
DDB_G0285871		25.926	8.228	47.089	0.311	1.411	RAN guanine nucleotide release factor
DDB_G0287763		10.933	3.392	7.010	0.310	0.508	RAG1AP1 family protein
DDB_G0268948		254.288	80.440	715.990	0.310	2.179	putative SAM dependent methyltransferase

DDB_G0277761		780.092	246.035	452.460	0.310	0.451	endoribonuclease L-PSP
DDB_G0277811	*cfaB*	146.023	44.940	102.845	0.309	0.559	counting factor associated protein
DDB_G0278701	*ncapD3*	3.029	0.967	10.317	0.308	2.600	non-SMC condensin II complex, subunit D3
DDB_G0272756	*csbC*	135.676	41.867	46.832	0.308	0.273	contact site B protein
DDB_G0293436	*abcA3*	7.861	2.481	20.738	0.308	2.029	ABC transporter A family protein
DDB_G0275119	*alyB*	202.889	63.092	87.781	0.307	0.339	lysozyme
DDB_G0278913		3.953	1.222	1.718	0.307	0.344	
DDB_G0270658	*acbA*	1224.544	375.688	1606.296	0.306	1.039	acyl-CoA binding protein
DDB_G0287125	*proA*	5384.775	1670.867	3077.692	0.306	0.446	profilin I
DDB_G0294058	*mrps13*	315.996	100.053	450.035	0.305	1.091	ribosomal protein S13, mitochondrial
DDB_G0275953		49.625	15.476	36.312	0.305	0.567	molybdenum cofactor sulfurase domain-containing protein
DDB_G0269892		178.969	55.131	156.022	0.305	0.681	
DDB_G0294070	*DidioMp26*	73.000	22.508	87.703	0.304	0.936	ribosomal protein S3, N-terminal domain-containing protein
DDB_G0288757		7.705	2.378	12.008	0.303	1.218	nicotinamide-nucleotide adenylyltransferase
DDB_G0277685		24.833	7.604	15.123	0.303	0.477	
DDB_G0280565		19.345	5.982	14.546	0.302	0.581	calponin homology (CH) domain-containing protein
DDB_G0274273		98.520	30.136	78.161	0.302	0.622	Double C2-like domain-containing protein beta
DDB_G0269200	*cplA*	101.806	31.022	155.122	0.302	1.197	calpain-like cysteine protease
DDB_G0285187		52.246	15.547	61.290	0.301	0.941	
DDB_G0272380		9.525	2.832	18.799	0.301	1.570	
DDB_G0288095		44.243	13.258	40.129	0.300	0.718	putative bactericidal permeability-increasing protein
DDB_G0287255	*gcvP*	45.902	14.079	43.777	0.299	0.738	glycine dehydrogenase (decarboxylating)
DDB_G0287507	*limD1*	106.833	32.235	167.931	0.299	1.239	LIM-type zinc finger-containing protein
DDB_G0291213		39.447	12.249	13.979	0.299	0.269	Skipper RT-IN
DDB_G0285825		19.216	5.694	16.981	0.298	0.702	
DDB_G0274669		56.114	16.637	55.094	0.297	0.780	
DDB_G0281495		22.031	6.502	33.244	0.297	1.201	
DDB_G0288759	*tmem14A*	231.374	69.727	284.619	0.296	0.957	transmembrane protein 14 A
DDB_G0280977	*abcC10*	2.885	0.863	7.607	0.295	2.067	ABC transporter C family protein
DDB_G0277773		4.528	1.362	2.246	0.294	0.384	solute carrier family 2 member protein
DDB_G0288333	*purB*	72.229	21.560	222.701	0.293	2.395	adenylosuccinate lyase
DDB_G0277453		244.438	72.343	96.443	0.292	0.308	
DDB_G0277129		4.136	1.253	8.741	0.292	1.608	
DDB_G0278873	*gnpda1*	112.416	33.174	46.281	0.292	0.323	glucosamine-6-phosphate deaminase
DDB_G0274137	*act14*	84.851	24.865	79.633	0.292	0.743	actin
DDB_G0279879		4.737	1.372	13.770	0.292	2.313	
DDB_G0278211		1.228	0.359	1.840	0.291	1.175	
DDB_G0278667		120.741	34.954	94.125	0.290	0.620	
DDB_G0293656	*fhkA*	1.695	0.511	3.166	0.289	1.430	RAD53 family protein kinase

DDB_G0294072	*mrps7*	111.323	32.955	126.461	0.289	0.881	ribosomal protein S7, mitochondrial
DDB_G0278341		438.196	128.415	466.389	0.289	0.833	putative ATP citrate lyase
DDB_G0274341		14.510	4.215	36.726	0.289	1.983	
DDB_G0286207		24.617	7.215	32.384	0.288	1.028	
DDB_G0284809	*isca2*	13.703	3.943	42.995	0.288	2.471	iron-sulfur cluster assembly 2 homolog
DDB_G0270646		2.513	0.715	4.034	0.288	1.286	
DDB_G0272320	*snrpb*	99.742	28.838	87.708	0.286	0.693	LSM (like-Sm) domain-containing protein
DDB_G0277183	*H4a*	823.084	237.658	208.095	0.286	0.199	histone H4
DDB_G0277749		6.007	1.774	2.701	0.285	0.344	
DDB_G0286387		3.420	0.983	11.865	0.285	2.708	meprin and TRAF homology (MATH) domain-containing protein
DDB_G0272965		7.626	2.170	12.394	0.284	1.287	
DDB_G0282973		9.611	2.703	7.837	0.284	0.653	
DDB_G0272548		100.127	29.329	278.190	0.284	2.124	TRE5-A ORF1
DDB_G0294068	*DidioMp27*	163.466	47.502	231.353	0.283	1.095	ribosomal protein S3, C-terminal domain-containing protein
DDB_G0279513		11.601	3.312	20.965	0.283	1.421	
DDB_G0293394	*pter*	6.249	1.766	10.962	0.283	1.390	phosphotriesterase-related protein
DDB_G0292224	*ecmG*	2.828	0.810	2.864	0.282	0.790	cellulose-binding domain-containing protein
DDB_G0274183		35.921	10.370	38.339	0.282	0.824	leucine-rich repeat-containing protein (LRR)
DDB_G0277615		334.547	95.291	209.440	0.281	0.491	ADF/cofilin-like domain-containing protein
DDB_G0272184		64.450	18.260	89.798	0.281	1.099	zinc-containing alcohol dehydrogenase (ADH)
DDB_G0346833		0.808	0.226	1.120	0.281	1.102	
DDB_G0294597		60.572	17.124	59.211	0.281	0.771	beta-lactamase family protein
DDB_G0294036	*mrpl2*	136.164	39.205	156.200	0.280	0.887	ribosomal protein L2, mitochondrial
DDB_G0276433		32.742	9.237	17.337	0.280	0.416	
DDB_G0270314	*crip*	129.830	36.283	132.715	0.280	0.816	LIM-type zinc finger-containing protein
DDB_G0278101	*cnrO*	30.783	8.778	28.142	0.279	0.710	putative cell number regulator
DDB_G0279081	*abpB*	1255.986	351.212	2005.027	0.279	1.266	34 kDa actin-binding protein
DDB_G0278941		3.294	0.941	3.905	0.279	0.910	
DDB_G0276925		28.398	7.912	34.587	0.278	0.965	
DDB_G0278725	*p17*	5104.296	1434.972	5411.781	0.278	0.831	
DDB_G0284461		6.968	1.937	16.851	0.277	1.910	leucine-rich repeat-containing protein (LRR)
DDB_G0277373		2.004	0.557	2.551	0.277	1.001	
DDB_G0278623		111.062	31.055	38.587	0.277	0.271	
DDB_G0268264		781.034	216.394	827.873	0.276	0.841	
DDB_G0294144		13.734	3.923	4.515	0.276	0.251	Skipper GAG-PRO
DDB_G0292606		7.526	2.070	13.267	0.276	1.398	
DDB_G0285519		7.854	2.141	10.340	0.276	1.050	Androgen-induced gene 1 protein
DDB_G0286341		60.584	16.539	164.143	0.275	2.162	putative glutathione transferase

DDB_G0294044	*mrps19*	162.692	45.781	176.052	0.275	0.841	ribosomal protein S19, mitochondrial
DDB_G0292016		25.623	7.039	45.126	0.274	1.398	von Willebrand factor A domain-containing protein 5A
DDB_G0267306		25.528	7.251	7.586	0.274	0.226	Skipper GAG-PRO-POL
DDB_G0293630		4.523	1.281	4.017	0.274	0.680	TRE5-B ORF2
DDB_G0286323		0.955	0.260	1.691	0.273	1.407	
DDB_G0291293	*rabU*	8.480	2.355	12.174	0.273	1.119	Rab GTPase
DDB_G0278731	*pgtD*	0.363	0.101	0.547	0.272	1.159	GPS (GPCR Proteolytic Site) domain-containing protein
DDB_G0275025		38.026	10.342	51.594	0.271	1.069	thioredoxin domain-containing protein
DDB_G0268212		5.575	1.482	47.762	0.270	6.861	putative DEAD/DEAH box helicase
DDB_G0285279		19.835	5.418	27.200	0.270	1.074	
DDB_G0287611		1.941	0.535	4.745	0.270	1.888	
DDB_G0271034		1.895	0.510	2.088	0.269	0.869	
DDB_G0293396	*gxcO*	2.227	0.614	7.580	0.269	2.628	RhoGEF domain-containing protein
DDB_G0271472		1.602	0.434	3.969	0.269	1.954	
DDB_G0282517		150.336	40.144	103.508	0.268	0.548	AhpC/TSA family protein
DDB_G0292880		6.201	1.703	15.501	0.267	1.924	putative protein tyrosine phosphatase, dual specificity
DDB_G0270924		22.247	6.050	31.027	0.266	1.075	
DDB_G0269910		61.925	16.680	79.203	0.265	0.997	SH3 domain-containing protein
DDB_G0290975		58.619	16.217	278.046	0.263	3.605	alpha/beta hydrolase fold-3 domain-containing protein
DDB_G0287067		26.545	7.062	31.362	0.262	0.922	
DDB_G0270048		1.162	0.314	2.480	0.262	1.643	
DDB_G0268590		1.574	0.415	11.340	0.261	5.632	
DDB_G0267958		8.208	2.217	21.099	0.261	1.970	SMAD/FHA domain-containing protein
DDB_G0284263		16.297	4.324	28.731	0.261	1.374	DNAJ heat shock N-terminal domain-containing protein
DDB_G0277777		12.813	3.380	9.896	0.261	0.604	
DDB_G0282603		16.563	4.379	16.548	0.261	0.781	NUDIX hydrolase family protein
DDB_G0272562	*csbB*	58.377	14.806	49.326	0.260	0.680	contact site B protein
DDB_G0293008	*cpnA*	42.232	11.110	51.875	0.260	0.965	copine A
DDB_G0285545		9.964	2.534	15.410	0.259	1.239	Ctr copper transporter family protein
DDB_G0293416	*abcB1*	25.782	6.839	70.171	0.259	2.098	ABC transporter B family protein
DDB_G0280621	*arrJ*	2.739	0.737	3.615	0.258	1.004	ADP-ribosylation factor-related
DDB_G0274203		75.769	19.471	62.511	0.258	0.658	
DDB_G0277417		2.454	0.639	1.621	0.258	0.519	
DDB_G0277455	*plbE*	209.563	53.901	75.982	0.258	0.288	phospholipase B-like protein
DDB_G0274601	*act9*	47.765	12.207	70.845	0.256	1.184	actin
DDB_G0276939	*pspB*	5.175	1.349	1.427	0.256	0.215	prespore protein
DDB_G0277429		17.696	4.630	15.737	0.256	0.690	BTB/POZ domain-containing protein
DDB_G0280543	*lmcA*	18.799	4.802	26.729	0.255	1.124	
DDB_G0276759	*cbpA*	4.831	1.230	9.847	0.255	1.612	calcium-binding protein

DDB_G0269684		2.539	0.645	21.674	0.254	6.732	
DDB_G0283449		92.068	22.927	49.847	0.254	0.438	
DDB_G0270192		9.087	2.298	11.910	0.253	1.040	DUF1624 family protein
DDB_G0292366		16.400	4.226	5.242	0.253	0.249	
DDB_G0279551		34.021	8.783	8.211	0.253	0.188	
DDB_G0285577		1.594	0.405	2.819	0.253	1.385	UPF0066 protein yaeB
DDB_G0269562		3.690	0.920	3.179	0.252	0.693	
DDB_G0272210	*idhM*	570.717	145.554	377.734	0.251	0.520	isocitrate dehydrogenase (NADP+), mitochondrial
DDB_G0283523		3.931	0.993	8.162	0.251	1.627	
DDB_G0272865		4.646	1.204	4.240	0.251	0.698	
DDB_G0288649	*ponL*	30.584	7.792	14.678	0.251	0.375	ponticulin-related protein
DDB_G0274575	*myoK*	27.495	6.865	30.145	0.250	0.870	myosin 1K heavy chain
DDB_G0287973		22.990	5.909	52.046	0.249	1.744	
DDB_G0269564		24.595	6.027	19.577	0.248	0.638	
DDB_G0280021	*mcfL*	21.699	5.479	26.817	0.247	0.961	mitochondrial substrate carrier family protein
DDB_G0289015		19.231	4.647	22.299	0.247	0.935	
DDB_G0267834		5.307	1.279	10.599	0.245	1.601	
DDB_G0287807		4.344	1.037	3.352	0.244	0.622	EGF-like domain-containing protein
DDB_G0280445		275.785	69.561	219.688	0.244	0.608	glutathione-dependent formaldehyde-activating, GFA family protein
DDB_G0290665	*gtaX*	0.907	0.226	6.263	0.243	5.333	GATA zinc finger domain-containing protein 24
DDB_G0277359		2.896	0.702	2.062	0.241	0.556	Pseudouridylate synthase PUS5
DDB_G0270276		6.006	1.483	98.222	0.240	12.554	
DDB_G0268892		18.027	4.341	211.047	0.239	9.233	
DDB_G0288931		18.372	4.340	23.501	0.238	1.019	
DDB_G0291394		54.841	13.153	62.417	0.238	0.898	Putative lipase YJR107W
DDB_G0279475	*gna1*	6.905	1.632	16.084	0.238	1.848	glucosamine 6-phosphate N-acetyltransferase
DDB_G0274291		21.314	4.970	75.947	0.237	2.857	putative T4-like lysozyme
DDB_G0289905		1.092	0.257	3.966	0.237	2.877	
DDB_G0276709		8.315	2.008	6.777	0.235	0.627	
DDB_G0269418		24.303	5.755	29.428	0.235	0.946	osmotically inducible family protein
DDB_G0288577		27.160	6.315	19.975	0.233	0.585	
DDB_G0271794	*captA*	6.329	1.466	11.166	0.231	1.398	CDP-alcohol phosphatidyltransferase
DDB_G0277855	*fimA*	192.785	44.989	221.433	0.231	0.905	fimbrin-1
DDB_G0278833		326.384	74.254	152.219	0.230	0.374	
DDB_G0277597	*gpt3*	37.681	8.876	7.308	0.229	0.149	putative glycophosphotransferase
DDB_G0293198	*kif10*	0.480	0.113	2.839	0.229	4.523	kinesin family member 10
DDB_G0284777		16.089	3.654	30.404	0.227	1.496	DUF1183 family protein
DDB_G0275989		2.190	0.499	3.909	0.226	1.402	beta-ketoacyl synthase family protein

DDB_G0271042		0.590	0.134	0.194	0.226	0.260	
DDB_G0291800		48.892	11.448	23.506	0.226	0.365	glutathione-dependent formaldehyde-activating, GFA family protein
DDB_G0281613		1.733	0.392	3.941	0.226	1.790	
DDB_G0269488		70.815	16.175	37.869	0.226	0.419	
DDB_G0292636		44.651	10.469	68.229	0.225	1.162	
DDB_G0291676		1.659	0.378	4.275	0.224	2.000	short-chain dehydrogenase/reductase (SDR) family protein
DDB_G0275243		9.101	2.079	12.280	0.224	1.047	PLAC8 family protein
DDB_G0274259		3.611	0.810	12.525	0.223	2.729	
DDB_G0284491		1.033	0.233	0.853	0.223	0.643	putative protein kinase
DDB_G0281065		12.657	2.781	10.164	0.222	0.644	
DDB_G0285869		3.329	0.752	22.549	0.222	5.272	
DDB_G0289497		19.474	4.319	19.914	0.222	0.810	
DDB_G0292136	*hatC*	9.016	1.956	14.593	0.222	1.309	hisactophilin III
DDB_G0289117	*myoB*	34.311	7.694	30.245	0.221	0.691	myosin IB
DDB_G0285101	*kif4*	1.300	0.292	5.413	0.221	3.252	kinesin-7
DDB_G0277495		3.912	0.854	3.639	0.221	0.751	
DDB_G0293162		1.602	0.359	5.294	0.221	2.569	
DDB_G0290411		7.666	1.728	30.031	0.221	3.065	
DDB_G0291532		7.789	1.687	12.537	0.220	1.285	
DDB_G0272242	*cupI*	9.354	2.132	59.181	0.220	4.822	cup family protein
DDB_G0291518		36.403	7.938	99.369	0.219	2.179	putative transmembrane protein
DDB_G0268144	*vwkA*	62.787	13.552	230.884	0.219	2.967	type A von Willebrand factor (VWFA) domain-containing protein
DDB_G0278073		2.509	0.563	10.323	0.219	3.167	
DDB_G0289663	*act5*	166.076	36.749	176.686	0.219	0.836	actin
DDB_G0270586		0.897	0.193	16.212	0.219	14.408	
DDB_G0286041	*iliJ*	1.766	0.383	9.204	0.218	4.151	
DDB_G0291834	*cpiA*	971.117	213.990	873.427	0.218	0.704	cysteine protease inhibitor
DDB_G0277189		44.329	9.662	39.334	0.217	0.701	
DDB_G0292434		0.972	0.213	4.737	0.216	3.804	
DDB_G0277475		0.747	0.159	2.966	0.216	3.154	
DDB_G0270306		1.216	0.257	1.597	0.216	1.055	putative transcriptional regulator
DDB_G0275125	*plbF*	124.946	27.285	95.017	0.215	0.593	phospholipase B-like protein
DDB_G0267308		11.628	2.583	4.363	0.213	0.282	Skipper GAG-PRO
DDB_G0291546		5.872	1.288	7.308	0.212	0.954	
DDB_G0280545	*act7*	191.315	40.849	258.457	0.212	1.067	actin
DDB_G0274655		8.037	1.725	49.164	0.212	4.785	
DDB_G0267752		0.427	0.089	2.026	0.211	3.791	Anoctamin-5

DDB_G0282255		441.571	92.585	1612.905	0.210	2.889	CMP/dCMP deaminase, zinc-binding domain-containing protein
DDB_G0277473		58.156	12.448	16.036	0.209	0.213	
DDB_G0292574		2.296	0.482	5.913	0.209	2.036	glucose/ribitol dehydrogenase family protein
DDB_G0287609		11.630	2.537	41.911	0.209	2.750	alpha/beta hydrolase fold-3 domain-containing protein
DDB_G0291255	*29C*	6717.473	1416.823	5200.335	0.209	0.608	
DDB_G0276893	*ctxB*	385.201	81.527	476.176	0.209	0.968	cortexillin II
DDB_G0293650	*fahd2*	16.013	3.357	22.454	0.208	1.103	Fumarylacetoacetate (FAA) hydrolase domain-containing protein
DDB_G0285183		1.222	0.258	6.355	0.208	4.035	
DDB_G0275293	*p2xB*	1.221	0.255	1.701	0.208	1.089	ATP-gated ion channel P2XB
DDB_G0274599	*act13*	126.356	26.167	172.334	0.208	1.088	actin
DDB_G0286465		2.817	0.609	10.639	0.207	2.865	putative protein kinase
DDB_G0274429		1.311	0.271	1.494	0.207	0.903	
DDB_G0274099	*cbpE*	18.734	3.882	54.989	0.206	2.308	calcium-binding protein
DDB_G0287047		6.073	1.262	16.970	0.206	2.199	
DDB_G0304886		22.397	4.571	21.006	0.205	0.746	LIM-type zinc finger-containing protein
DDB_G0271548		13.530	2.809	21.249	0.204	1.225	
DDB_G0270126	*rasY*	2.593	0.548	3.648	0.204	1.082	Ras GTPase
DDB_G0286307		25.011	5.162	25.692	0.204	0.802	
DDB_G0269548	*limG*	3.404	0.686	7.658	0.204	1.796	LIM-type zinc finger-containing protein
DDB_G0288591		31.564	6.611	30.262	0.203	0.736	short-chain dehydrogenase/reductase (SDR) family protein
DDB_G0283021	*sodB*	28.939	5.752	17.900	0.202	0.499	superoxide dismutase
DDB_G0291169		0.841	0.167	0.860	0.201	0.812	
DDB_G0267432	*abcG15*	110.262	22.640	184.886	0.200	1.289	ABC transporter G family protein
DDB_G0285215		114.049	22.708	141.560	0.200	0.985	
DDB_G0281219		70.737	14.489	98.483	0.199	1.070	Nucleoside diphosphate-linked moiety X motif 6
DDB_G0270078		1.656	0.332	1.125	0.197	0.526	
DDB_G0287831		3.084	0.625	4.944	0.197	1.233	
DDB_G0276383		15.433	3.123	37.262	0.197	1.868	putative guanylate cyclase
DDB_G0289391	*pkiA*	2145.114	426.783	3088.450	0.196	1.125	putative protein kinase C inhibitor
DDB_G0271468		0.268	0.054	0.418	0.195	1.187	
DDB_G0288697		17.801	3.626	37.670	0.195	1.612	
DDB_G0274181		1062.864	214.685	798.257	0.195	0.573	glycoside hydrolase family 25 protein
DDB_G0286351		0.590	0.115	3.514	0.194	4.670	putative transcriptional regulator
DDB_G0276445		228.115	44.840	659.365	0.194	2.265	heat shock protein Hsp70 family protein
DDB_G0290177		174.425	33.560	464.016	0.194	2.117	putative transmembrane protein
DDB_G0276451		3.970	0.781	10.567	0.193	2.069	
DDB_G0270002		0.998	0.194	3.427	0.193	2.692	

DDB_G0277405		261.522	51.827	280.003	0.193	0.829	SAP DNA-binding domain-containing protein
DDB_G0269592		2.896	0.553	7.897	0.192	2.174	
DDB_G0283101		6.742	1.266	44.231	0.187	5.181	
DDB_G0281551	*guaA*	60.150	11.569	410.262	0.187	5.227	GMP synthetase
DDB_G0287607	*pcna*	55.884	10.459	225.141	0.187	3.189	proliferating cell nuclear antigen
DDB_G0277807		98.449	18.326	46.541	0.187	0.377	
DDB_G0269248	*cf45-1*	388.131	73.948	277.045	0.186	0.551	component of the counting factor (CF) complex
DDB_G0284631		9.777	1.811	31.243	0.186	2.550	
DDB_G0276049		551.341	102.747	530.528	0.185	0.757	
DDB_G0295701		1.375	0.251	18.850	0.184	11.018	paramecium surface antigen repeat-containing protein
DDB_G0267742		0.347	0.064	0.660	0.184	1.509	putative acylaminoacyl-peptidase
DDB_G0280011		17.999	3.252	9.101	0.184	0.405	transposable element
DDB_G0270566		4.674	0.878	28.175	0.183	4.639	DDHD domain-containing protein
DDB_G0278111		70.772	13.217	62.931	0.183	0.692	Programmed cell death protein 5
DDB_G0267420	*sodA*	1133.538	210.441	1163.788	0.183	0.801	superoxide dismutase
DDB_G0279583		0.477	0.086	0.951	0.182	1.573	
DDB_G0284645		47.652	8.478	46.498	0.181	0.788	
DDB_G0287899	*rif1*	1.033	0.195	3.488	0.180	2.556	
DDB_G0293940		1.687	0.308	2.441	0.180	1.120	
DDB_G0270122	*rasW*	4.276	0.763	13.694	0.180	2.578	Ras GTPase
DDB_G0274987		54.459	9.813	49.401	0.180	0.717	
DDB_G0270124	*rasX*	20.392	3.847	11.585	0.180	0.431	Ras GTPase
DDB_G0277357		25.656	4.839	12.653	0.179	0.371	RmlC-like cupin family protein
DDB_G0281767		0.996	0.180	0.780	0.179	0.613	BTB/POZ domain-containing protein
DDB_G0286431		1.009	0.178	26.809	0.179	21.327	
DDB_G0290793	*rabN2*	1.596	0.296	2.080	0.178	0.997	Rab GTPase
DDB_G0349136		1.798	0.317	2.689	0.178	1.187	
DDB_G0279707	*iliP*	10.520	1.845	46.981	0.177	3.571	DUF3430 family protein
DDB_G0280335		98.508	18.038	59.247	0.176	0.455	
DDB_G0268026		166.749	29.303	236.116	0.176	1.126	transmembrane protein
DDB_G0274433		0.712	0.127	0.735	0.175	0.800	CBS (cystathionine-beta-synthase) domain-containing protein
DDB_G0276043	*ifkC*	5.850	1.008	9.040	0.175	1.244	PEK family protein kinase
DDB_G0291416		7.894	1.389	8.551	0.175	0.853	
DDB_G0274321	*purF*	64.237	11.403	251.746	0.174	3.047	amidophosphoribosyltransferase
DDB_G0286333		0.360	0.064	3.206	0.173	6.833	
DDB_G0279851		23.426	4.232	28.816	0.173	0.928	GNAT family protein
DDB_G0277083	*uduD1*	28.451	4.866	49.939	0.173	1.405	N-terminal delta endotoxin domain-containing protein
DDB_G0271286	*clasp*	0.886	0.155	4.677	0.173	4.125	CLIP-associating protein

DDB_G0288919	*psiI*	0.694	0.118	0.717	0.172	0.831	PA14 domain-containing protein
DDB_G0292312		7.120	1.242	9.353	0.172	1.034	
DDB_G0291592		2.239	0.396	3.410	0.171	1.161	
DDB_G0281071	*serA*	188.050	32.417	387.640	0.170	1.614	3-phosphoglycerate dehydrogenase
DDB_G0279877		7.573	1.307	14.929	0.170	1.538	
DDB_G0278009	*racN*	1.873	0.314	1.916	0.169	0.812	Rho GTPase
DDB_G0290325	*espl1*	1.127	0.194	2.945	0.168	2.009	separase
DDB_G0274359	*cbpK*	8.554	1.440	19.534	0.167	1.790	NCS-1/frequenin-related protein
DDB_G0295807		10.333	1.774	17.015	0.167	1.264	short-chain dehydrogenase/reductase (SDR) family protein
DDB_G0271666	*prtB*	34.385	5.655	77.930	0.165	1.801	proteosomal alpha-subunit 7-1
DDB_G0283309		115.204	19.042	109.834	0.165	0.754	translation initiation factor eIF-2B alpha subunit
DDB_G0274129	*act12*	14.134	2.366	19.698	0.164	1.087	actin
DDB_G0291472		7.200	1.179	7.206	0.164	0.793	pirin family protein
DDB_G0274143		1.750	0.292	5.725	0.163	2.516	
DDB_G0272466		8.006	1.310	84.149	0.162	8.163	short-chain dehydrogenase/reductase (SDR) family protein
DDB_G0282419	*cyp508B1*	1.238	0.198	5.691	0.162	3.682	cytochrome P450 family protein
DDB_G0277291		0.548	0.086	1.175	0.161	1.751	EGF-like domain-containing protein
DDB_G0284657		7.736	1.261	15.134	0.161	1.522	Tripartite motif-containing protein 58
DDB_G0288819	*ncapG2*	1.131	0.184	5.301	0.161	3.643	non-SMC condensin II complex, subunit G2
DDB_G0274223		7.914	1.311	15.030	0.160	1.457	putative glutathione S-transferase
DDB_G0274115	*abcG12*	59.193	9.669	244.392	0.160	3.189	ABC transporter G family protein
DDB_G0267438	*abcA2*	17.473	2.840	29.302	0.160	1.308	ABC transporter A family protein
DDB_G0281073		1.222	0.196	2.401	0.159	1.538	
DDB_G0291806		1.759	0.288	8.664	0.158	3.773	
DDB_G0289863		0.410	0.064	4.735	0.157	9.198	
DDB_G0272238		570.898	87.051	336.755	0.157	0.481	
DDB_G0269214	*abcG1*	39.325	6.360	115.440	0.157	2.258	ABC transporter G family protein
DDB_G0290001		74.168	11.948	26.335	0.157	0.274	
DDB_G0282041		4.241	0.678	5.608	0.157	1.025	
DDB_G0278477	*sarB*	10.453	1.636	17.266	0.156	1.306	ARF/SAR superfamily protein
DDB_G0282033		24.506	3.985	37.435	0.156	1.165	major vault protein
DDB_G0305714		28.637	4.568	4.771	0.155	0.128	Skipper ORF2
DDB_G0275439	*cad2*	19.613	3.139	9.098	0.154	0.355	putative adhesion molecule
DDB_G0280201		18.507	2.757	636.700	0.153	27.815	
DDB_G0289549		26.664	4.062	8.278	0.152	0.245	
DDB_G0277097	*uduF*	42.971	6.329	183.249	0.150	3.440	
DDB_G0284137		1.702	0.264	3.097	0.150	1.396	
DDB_G0277781		19.297	2.968	4.240	0.150	0.169	Skipper GAG-PRO
DDB_G0280045	*thyA*	21.211	3.157	259.467	0.148	9.573	thymidylate synthase (FAD)

DDB_G0286367		0.815	0.119	2.099	0.147	2.041	
DDB_G0283911		67.089	9.591	157.988	0.145	1.897	HSP20-like chaperone domain-containing protein
DDB_G0272094		16.925	2.400	47.033	0.145	2.234	
DDB_G0274523		0.781	0.115	1.186	0.145	1.183	RNA recognition motif-containing protein RRM
DDB_G0289811	*act10*	161.175	23.387	301.726	0.144	1.483	actin
DDB_G0269788		102.697	14.550	67.593	0.144	0.531	methylsterol monooxygenase
DDB_G0274209		0.682	0.097	1.133	0.144	1.320	
DDB_G0291646		500.799	70.650	623.830	0.143	0.996	
DDB_G0289869		2.860	0.432	2.863	0.143	0.758	
DDB_G0279343	*aurK*	5.841	0.835	35.317	0.143	4.783	Aurora family protein kinase
DDB_G0286059	*asns*	102.401	14.745	149.193	0.143	1.146	asparagine synthetase
DDB_G0277489		30.617	4.348	910.948	0.143	23.811	
DDB_G0268816		21.898	3.262	20.038	0.143	0.692	
DDB_G0270700		0.795	0.116	3.567	0.142	3.480	calcium-binding EGF domain-containing protein
DDB_G0286363		11.032	1.630	13.663	0.141	0.935	
DDB_G0288579		250.527	35.289	162.793	0.141	0.515	
DDB_G0289575		1.432	0.207	0.742	0.140	0.393	
DDB_G0290659	*sdrA*	82.742	11.833	200.315	0.139	1.864	short-chain dehydrogenase/reductase (SDR) family protein
DDB_G0289533	*act27*	3.568	0.522	16.545	0.139	3.483	actin
DDB_G0269520	*abnC*	74.918	10.747	56.672	0.138	0.584	actobindin
DDB_G0281049		87.017	12.051	136.115	0.138	1.247	
DDB_G0294246		2.799	0.394	0.781	0.138	0.215	Tdd-4
DDB_G0267310		102.845	14.093	18.789	0.134	0.141	Skipper GAG-PRO
DDB_G0276563		17.165	2.297	21.475	0.134	0.992	Four and a half LIM domains protein 1
DDB_G0284629		140.155	19.077	241.325	0.133	1.341	
DDB_G0288145	*purL*	71.082	9.590	299.776	0.133	3.286	phosphoribosylformylglycinamide synthase
DDB_G0288631		1.635	0.213	4.711	0.132	2.308	peptidase S9 family protein
DDB_G0281695		7.536	1.022	17.357	0.131	1.762	
DDB_G0280501		0.895	0.116	3.143	0.130	2.801	
DDB_G0278699		12.498	1.658	42.830	0.130	2.701	
DDB_G0277355		494.459	66.258	438.893	0.130	0.679	
DDB_G0279591	*glnA3*	211.051	27.795	740.394	0.130	2.725	glutamine synthetase type III
DDB_G0275519		104.149	13.808	94.371	0.129	0.701	
DDB_G0283097		2.859	0.358	9.970	0.127	2.772	
DDB_G0285697		0.509	0.065	0.697	0.126	1.074	
DDB_G0278345	*acly*	414.686	52.680	744.526	0.126	1.416	ATP citrate lyase
DDB_G0278593	*exdl2B*	4.307	0.550	5.141	0.125	0.926	3'-5' exonuclease domain-containing protein
DDB_G0277421		4.410	0.556	1.846	0.125	0.327	
DDB_G0289731		0.668	0.085	7.334	0.125	8.556	N-terminal delta endotoxin domain-containing protein

DDB_G0289271	*gins1*	1.453	0.183	5.554	0.124	2.992	GINS complex subunit 1
DDB_G0277227		95.219	11.555	72.498	0.123	0.608	
DDB_G0267370		8.835	1.091	6.789	0.121	0.594	DDT-A
DDB_G0282209		51.376	6.443	72.373	0.120	1.073	type A von Willebrand factor (VWFA) domain-containing protein
DDB_G0287465		1.714	0.209	2.567	0.120	1.169	
DDB_G0268064		152.870	18.553	235.867	0.120	1.207	esterase/lipase/thioesterase domain-containing protein
DDB_G0293648		1.971	0.238	1.277	0.119	0.504	
DDB_G0288395		5.241	0.619	7.613	0.119	1.150	
DDB_G0267786		28.836	3.439	34.751	0.118	0.957	pleckstrin homology (PH) domain-containing protein
DDB_G0269838		2.905	0.345	4.851	0.118	1.304	
DDB_G0271140	*abcA7*	7.294	0.846	14.390	0.118	1.586	ABC transporter A family protein
DDB_G0291408		17.157	2.109	45.041	0.117	1.988	
DDB_G0284549	*dduF*	18.330	2.075	337.066	0.117	14.912	speract/scavenger receptor domain-containing protein
DDB_G0275001		35.574	4.259	26.928	0.117	0.586	UPF0244 protein yjjX
DDB_G0276277		30.023	3.397	38.696	0.116	1.057	
DDB_G0278171		2.741	0.323	3.043	0.116	0.867	
DDB_G0268344		12.502	1.455	12.799	0.116	0.809	
DDB_G0293460		103.168	11.736	153.580	0.115	1.187	
DDB_G0272849		6.730	0.772	10.516	0.115	1.238	EGF-like domain-containing protein
DDB_G0291572		5.128	0.574	3.664	0.114	0.579	
DDB_G0279901		6.067	0.678	38.274	0.114	5.048	
DDB_G0268412		2.867	0.332	9.724	0.113	2.622	
DDB_G0291758		4.272	0.488	11.145	0.113	2.030	
DDB_G0272688		0.574	0.066	1.516	0.113	2.071	NmrA-like protein
DDB_G0267592		278.872	31.433	19.457	0.112	0.055	
DDB_G0277253		5.687	0.630	7.120	0.110	0.988	MS ion channel domain-containing protein
DDB_G0272314		0.681	0.077	0.651	0.110	0.732	
DDB_G0274133	*act2*	24.501	2.722	31.637	0.110	1.014	actin
DDB_G0284931		499.835	54.949	43.043	0.109	0.068	
DDB_G0295739		55.665	6.216	179.040	0.108	2.477	
DDB_G0275743		1.276	0.140	3.473	0.108	2.137	
DDB_G0288935	*panC*	39.985	4.364	48.521	0.108	0.953	pantoate-beta-alanine ligase
DDB_G0279683		8.681	0.947	17.033	0.107	1.521	putative transmembrane protein
DDB_G0288921	*hspI*	7.582	0.840	32.479	0.107	3.252	heat shock protein Hsp20 domain-containing protein
DDB_G0276969		18.829	1.989	50.253	0.106	2.120	putative carboxylesterase
DDB_G0287161		15.276	1.638	27.386	0.106	1.407	
DDB_G0291740		44.513	4.661	171.937	0.106	3.100	
DDB_G0275341		1.188	0.125	3.319	0.105	2.194	

DDB_G0269180	*rcdP*	2.661	0.274	15.080	0.103	4.474	random cDNA clone veg114
DDB_G0285939	*ctnnA*	6.325	0.657	17.279	0.102	2.136	alpha-catenin related protein
DDB_G0291281		39.350	4.016	134.269	0.102	2.699	
DDB_G0277367		50.060	5.134	24.395	0.102	0.382	
DDB_G0286911		10.868	1.085	15.990	0.101	1.185	
DDB_G0292424		7.842	0.792	2.213	0.099	0.218	
DDB_G0285791		1.184	0.117	3.089	0.099	2.047	
DDB_G0283197		0.558	0.054	2.350	0.099	3.397	C2H2-type zinc finger-containing protein
DDB_G0276899		82.658	8.421	122.592	0.098	1.126	thioredoxin fold domain-containing protein
DDB_G0293722		0.655	0.063	0.136	0.096	0.165	
DDB_G0267352		0.190	0.018	0.024	0.095	0.098	DDT-A
DDB_G0283915		385.100	36.228	194.325	0.093	0.395	esterase/lipase/thioesterase domain-containing protein
DDB_G0291712		0.479	0.042	0.617	0.091	1.044	
DDB_G0290329		31.887	2.898	52.942	0.091	1.314	short-chain dehydrogenase/reductase (SDR) family protein
DDB_G0270022		7.248	0.655	22.615	0.091	2.468	cellular retinaldehyde-binding/triple function domain-containing protein
DDB_G0293640		3.362	0.304	0.872	0.090	0.208	
DDB_G0272034		0.331	0.031	0.846	0.090	1.931	
DDB_G0291137		17.188	1.622	3.416	0.090	0.149	Lysozyme-like protein 2
DDB_G0294567		0.955	0.086	4.956	0.089	4.045	
DDB_G0289097		1.380	0.122	1.254	0.089	0.730	
DDB_G0292028		4.232	0.386	22.005	0.088	3.983	von Willebrand factor A domain-containing protein 5B1
DDB_G0284791	*cpnC*	9.663	0.872	9.595	0.087	0.763	copine C
DDB_G0292528		23.387	2.010	107.225	0.087	3.676	pirin family protein
DDB_G0267688	*cpnD*	17.485	1.616	24.755	0.087	1.062	copine D
DDB_G0267734		498.854	42.896	646.197	0.085	1.020	putative SAM dependent methyltransferase
DDB_G0277827	*cafA*	61.636	5.211	101.031	0.085	1.308	calfumirin-1
DDB_G0270490		1.730	0.150	2.041	0.085	0.916	
DDB_G0269202	*gdcA*	32.983	2.801	200.578	0.085	4.826	gp64 and disintegrin-like, cysteine-rich protein
DDB_G0271480		2.106	0.178	2.725	0.084	1.024	
DDB_G0277261		2.597	0.217	1.976	0.083	0.601	
DDB_G0288745		6.404	0.536	6.432	0.083	0.789	
DDB_G0295669		16.615	1.335	10.256	0.082	0.499	isochorismatase hydrolase
DDB_G0292080		19.552	1.645	337.686	0.082	13.324	
DDB_G0267490	*ChLim*	7.378	0.602	17.608	0.082	1.902	calponin homology (CH) domain-containing protein
DDB_G0271032		2.616	0.214	0.131	0.081	0.039	EGF-like domain-containing protein
DDB_G0268862		29.599	2.404	171.989	0.081	4.575	major facilitator superfamily protein
DDB_G0280249		1.766	0.143	13.672	0.080	6.106	MAP65/ASE1 family protein
DDB_G0269902	*act26*	1.797	0.147	0.694	0.080	0.300	actin

DDB_G0271852	*rtaA*	0.873	0.068	3.929	0.078	3.534	lipid-translocating exporter family protein
DDB_G0280727		21.454	1.654	65.842	0.078	2.445	
DDB_G0282839		1.453	0.110	1.728	0.077	0.962	
DDB_G0269322		37.725	2.932	279.037	0.077	5.802	
DDB_G0286369		0.950	0.073	2.170	0.077	1.801	
DDB_G0267362		0.526	0.041	0.044	0.077	0.065	Skipper GAG-PRO-POL
DDB_G0276513		62.720	4.891	110.966	0.076	1.378	
DDB_G0292664		248.875	19.061	232.315	0.076	0.736	calponin homology (CH) domain-containing protein
DDB_G0277103		2.505	0.192	3.237	0.076	1.014	
DDB_G0285685		0.717	0.052	0.232	0.075	0.262	
DDB_G0277725	*thfA*	106.548	8.145	192.164	0.075	1.400	methylenetetrahydrofolate dehydrogenase (NAD+)
DDB_G0279999		0.736	0.054	1.409	0.075	1.539	EGF-like domain-containing protein
DDB_G0287765		36.992	2.807	8.141	0.074	0.171	NADH:flavin oxidoreductase/NADH oxidase domain-containing protein
DDB_G0292340		0.855	0.062	3.317	0.074	3.102	patatin family protein
DDB_G0277967		1.163	0.086	7.417	0.073	5.000	putative GTPase activating protein (GAP)
DDB_G0276795	*ctbsB*	30.416	2.186	25.045	0.073	0.661	glycoside hydrolase family 18 protein
DDB_G0286021		50.565	3.672	16.247	0.073	0.254	
DDB_G0269482		85.601	6.064	144.226	0.073	1.370	
DDB_G0277469		2.555	0.183	3.532	0.072	1.097	
DDB_G0288635		86.880	6.172	568.003	0.072	5.231	transmembrane protein
DDB_G0270134		23.165	1.661	40.385	0.072	1.377	
DDB_G0267756		8.807	0.609	20.384	0.071	1.888	
DDB_G0272506		11.568	0.775	43.565	0.070	3.078	
DDB_G0293370		16.985	1.176	16.000	0.069	0.751	
DDB_G0272368		0.854	0.059	6.905	0.069	6.375	
DDB_G0274165		0.793	0.053	2.098	0.068	2.138	
DDB_G0278825		18.578	1.250	75.941	0.068	3.267	
DDB_G0271696	*prtA*	1.447	0.097	5.677	0.068	3.112	proteosomal alpha-subunit M3
DDB_G0288143		9.518	0.615	23.087	0.067	1.984	lysozyme C family protein
DDB_G0286551	*gerE*	12.397	0.843	10.116	0.067	0.632	
DDB_G0275553		1.341	0.089	11.938	0.066	7.019	
DDB_G0280323		23.628	1.524	6.054	0.066	0.206	
DDB_G0291894	*ncapH2*	0.964	0.065	7.780	0.065	6.131	non-SMC condensin II complex, subunit H2
DDB_G0270922		285.226	18.201	729.908	0.064	2.033	
DDB_G0274895		12.861	0.842	20.706	0.063	1.245	
DDB_G0285127		1.081	0.067	3.444	0.063	2.576	
DDB_G0270726	*uduC*	76.290	4.687	37.121	0.063	0.395	pyridoxal phosphate-dependent decarboxylase family protein
DDB_G0294190		15.503	1.004	7.762	0.062	0.379	DDT-A

DDB_G0286317		0.833	0.051	1.252	0.061	1.196	
DDB_G0278817		1.200	0.073	1.739	0.061	1.135	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase-like protein 1
DDB_G0268318		2.747	0.163	17.475	0.061	5.144	
DDB_G0269354		4.633	0.275	7.108	0.061	1.246	EGF-like domain-containing protein
DDB_G0275033		2.890	0.167	11.501	0.060	3.261	
DDB_G0270088		13.057	0.743	20.236	0.058	1.252	
DDB_G0283913		10.083	0.561	51.489	0.058	4.161	heat shock protein Hsp20 domain-containing protein
DDB_G0278581	*psiF*	46.392	2.637	79.779	0.057	1.381	PA14 domain-containing protein
DDB_G0286953	*uduG*	364.525	21.305	328.824	0.057	0.697	frizzled cysteine-rich domain-containing protein
DDB_G0274501	*mcfY*	2.170	0.124	3.679	0.057	1.334	mitochondrial substrate carrier family protein
DDB_G0274499		39.164	2.191	23.492	0.057	0.484	patatin family protein
DDB_G0277693		0.785	0.045	2.985	0.057	2.979	
DDB_G0284437		1.316	0.075	2.155	0.055	1.250	DNA repair protein REV1
DDB_G0282167		0.842	0.045	0.720	0.054	0.686	
DDB_G0274773	*fslH*	9.178	0.489	16.002	0.054	1.392	G-protein-coupled receptor (GPCR) family protein
DDB_G0281117		10.551	0.550	55.013	0.054	4.202	
DDB_G0285203		13.178	0.694	30.385	0.053	1.841	
DDB_G0285003	*aifC*	1.138	0.061	13.651	0.053	9.392	putative apoptosis inducing factor
DDB_G0277707		5.492	0.283	2.390	0.052	0.350	
DDB_G0270100		0.949	0.049	1.544	0.051	1.264	
DDB_G0270430	*wipA*	134.745	6.640	81.878	0.050	0.494	WH2 domain-containing protein
DDB_G0277281		2.180	0.112	3.306	0.050	1.153	TRE5-A ORF1
DDB_G0277203		59.142	3.005	36.693	0.050	0.479	NAD-dependent epimerase/dehydratase family protein
DDB_G0278243	*rliB*	16.423	0.804	57.667	0.048	2.727	
DDB_G0277383		295.327	14.045	158.108	0.048	0.428	
DDB_G0277793		7.849	0.369	18.818	0.048	1.931	
DDB_G0290539		7.750	0.374	4.786	0.047	0.471	
DDB_G0282919		4.983	0.237	0.621	0.046	0.096	Tdd-4
DDB_G0277785		40.784	1.964	23.291	0.046	0.435	DDT-A
DDB_G0283429		5.272	0.236	19.067	0.046	2.930	
DDB_G0291071		1.287	0.058	7.016	0.045	4.267	
DDB_G0275487		1.421	0.063	4.200	0.045	2.363	
DDB_G0348143		29.784	1.290	9.623	0.045	0.263	
DDB_G0283643		20.530	0.869	18.194	0.044	0.723	
DDB_G0292004	*docC*	1.246	0.054	6.281	0.044	4.007	cyclin-like F-box containing protein
DDB_G0292368		15.230	0.669	16.039	0.043	0.822	
DDB_G0285245		16.483	0.675	98.268	0.042	4.826	
DDB_G0287305		11.919	0.493	1.616	0.042	0.108	

DDB_G0290637		5.059	0.203	17.250	0.042	2.777	
DDB_G0281077		0.705	0.030	1.472	0.041	1.596	
DDB_G0270132		44.271	1.885	28.771	0.040	0.487	
DDB_G0291658		1.631	0.063	2.877	0.039	1.424	
DDB_G0267790		8.986	0.331	29.118	0.038	2.659	
DDB_G0275911		212.969	8.145	169.765	0.038	0.624	
DDB_G0282737		62.996	2.319	90.379	0.036	1.100	
DDB_G0288003		0.412	0.014	1.302	0.036	2.543	EGF-like domain-containing protein
DDB_G0272919		4.454	0.153	16.631	0.035	3.034	
DDB_G0267456	*cbp2*	287.385	10.031	491.490	0.035	1.340	calcium-binding protein
DDB_G0274245		2.549	0.091	6.247	0.034	1.866	
DDB_G0274485		28.254	0.939	49.659	0.034	1.436	
DDB_G0268258	*kif12*	1.956	0.064	13.091	0.032	5.257	kinesin family member 12
DDB_G0267354		52.425	1.733	21.579	0.032	0.314	DDT-A
DDB_G0281059		4.274	0.139	3.221	0.031	0.567	methyltransferase type 12 domain-containing protein
DDB_G0282559	*dduA*	46.376	1.409	226.411	0.031	3.901	metallophosphoesterase domain-containing protein
DDB_G0276097		9.035	0.264	32.030	0.030	2.884	putative transmembrane protein
DDB_G0286627		0.754	0.022	0.094	0.030	0.098	protein kinase, STE group
DDB_G0283271		195.245	5.746	75.685	0.029	0.307	cytochrome b561 / ferric reductase transmembrane domain-containing protein
DDB_G0270950		1.186	0.035	0.081	0.028	0.052	TRE5-B ORF2
DDB_G0291289		0.785	0.022	0.063	0.028	0.064	
DDB_G0271692	*comC*	1.914	0.053	3.123	0.028	1.296	EGF-like domain-containing protein
DDB_G0305372		11.318	0.320	11.222	0.028	0.772	
DDB_G0284925		56.193	1.594	7.692	0.028	0.107	
DDB_G0276597		3.630	0.100	5.363	0.028	1.160	
DDB_G0277599		88.047	2.546	17.159	0.027	0.144	
DDB_G0268352		1.704	0.045	3.577	0.026	1.648	EGF-like domain-containing protein
DDB_G0288461	*hgsA*	207.377	5.232	244.701	0.025	0.943	hydroxymethylglutaryl-CoA synthase
DDB_G0275475		15.184	0.373	57.030	0.024	2.823	
DDB_G0278503		25.693	0.606	40.218	0.023	1.236	
DDB_G0276219		3.293	0.074	25.185	0.023	6.211	putative transmembrane protein
DDB_G0277689		3.948	0.091	1.139	0.023	0.225	
DDB_G0290885		105.916	2.182	139.710	0.021	1.063	
DDB_G0291344		34.385	0.709	2.733	0.020	0.062	carboxylesterase, type B family protein
DDB_G0277333		161.850	3.300	99.866	0.020	0.469	
DDB_G0289629		16.057	0.262	37.432	0.016	1.804	
DDB_G0292334		22.873	0.283	36.219	0.013	1.279	
DDB_G0290993		39.818	0.468	1.809	0.012	0.037	

DDB_G0282815		5129.721	57.411	5409.709	0.011	0.821	
DDB_G0281535		4.141	0.042	0.853	0.010	0.159	Skipper GAG-PRO-POL
DDB_G0270058		4.718	0.048	1.342	0.009	0.206	
DDB_G0291229	*abpF*	97.625	0.476	1.151	0.005	0.009	

Table S3: Gene ontology (GO) term enrichment analysis of positively CbfA-regulated genes.

Biological Process					
GO Term	**Aspect**	**P-value**	**Sample frequency**	**Background frequency**	**Genes**
GO:0000902 cell morphogenesis	P	5.14e-03	11/269 (4.1%)	44/5436 (0.8%)	act27 act12 act9 ctxB act14 act7 act5 act10 act2 act26 act13
GO:0032989 cellular component morphogenesis	P	1.27e-02	11/269 (4.1%)	48/5436 (0.9%)	act27 act12 act9 ctxB act14 act7 act5 act10 act2 act26 act13
GO:0051301 cell division	P	1.32e-02	14/269 (5.2%)	76/5436 (1.4%)	act12 act9 vwkA ctxB act14 act7 act5 ctnnA kif12 act10 proA cpnA act2 act13
GO:0000910 cytokinesis	P	1.32e-02	14/269 (5.2%)	76/5436 (1.4%)	act12 act9 vwkA ctxB act14 act7 act5 ctnnA kif12 act10 proA cpnA act2 act13

Cellular Component					
GO Term	**Aspect**	**P-value**	**Sample frequency**	**Background frequency**	**Genes**
GO:0005856 cytoskeleton	C	5.51e-04	29/269 (10.8%)	217/5436 (4.0%)	DDB_G0277203 DDB_G0268212 act27 DDB_G0290593 act12 act9 DDB_G0276899 DDB_G0268064 myoB DDB_G0282815 act14 act7 act5 forG kif12 act10 ChLim proA DDB_G0277615 DDB_G0281613 act2 pkiA clasp aurK DDB_G0267734 DDB_G0268948 abpB act26 act13
GO:0001891 phagocytic cup	C	1.07e-03	10/269 (3.7%)	31/5436 (0.6%)	fimA act12 act9 act14 act7 act5 act10 ChLim act2 act13
GO:0045335 phagocytic vesicle	C	1.73e-03	35/269 (13.0%)	307/5436 (5.6%)	DDB_G0277203 sdrA fimA sodA DDB_G0290593 acly act12 DDB_G0278341 act9 DDB_G0283911 DDB_G0292574 DDB_G0268064 serA myoB asns ctxB act14 act7 act5 act10 DDB_G0282209 ChLim DDB_G0287609 DDB_G0277615 guaA cpnA act2 pkiA DDB_G0268948 glnA3 cpiA thyA abpB DDB_G0276445 act13
GO:0005884 actin filament	C	1.97e-03	8/269 (3.0%)	20/5436 (0.4%)	act12 act9 act14 act7 act5 act10 act2 act13
GO:0031252 cell leading edge	C	2.76e-03	13/269 (4.8%)	58/5436 (1.1%)	fimA act12 act9 myoB ctxB act14 act7 act5 act10 wipA act2 cbpA act13
GO:0030139 endocytic vesicle	C	3.15e-03	35/269 (13.0%)	315/5436 (5.8%)	DDB_G0277203 sdrA fimA sodA DDB_G0290593 acly act12 DDB_G0278341 act9 DDB_G0283911 DDB_G0292574 DDB_G0268064 serA myoB asns ctxB act14 act7 act5 act10 DDB_G0282209 ChLim DDB_G0287609 DDB_G0277615 guaA cpnA act2 pkiA DDB_G0268948 glnA3 cpiA thyA abpB DDB_G0276445 act13
GO:0031410 cytoplasmic vesicle	C	4.05e-03	38/269 (14.1%)	360/5436 (6.6%)	DDB_G0277203 sdrA fimA sodA DDB_G0290593 acly act12 DDB_G0278341 act9 ponB vwkA DDB_G0283911 DDB_G0292574 DDB_G0268064 serA myoB asns ctxB act14 act7 act5 act10 DDB_G0282209 ChLim DDB_G0287609

9.2 Ergänzungen Manuskript 3

The *Dictyostelium discoideum* RNA-dependent RNA polymerase RrpC silences the centromeric retrotransposon DIRS-1 post-transcriptionally and is required for the spreading of RNA silencing signals

---

Stephan Wiegand, Doreen Meier, Carsten Seehafer, Marek Malicki, Patrick Hofmann, **Anika Schmith**, Thomas Winckler, Balint Földesi, Benjamin Boesler, Wolfgang Nellen, Johan Reimegard, Max Käller, Jimmie Hällman, Olof Emanualsson, Lotta Avesson, Frederik Söderbom und Christian Hammann

Publiziert in: *Nucleic Acids Research*, 2014

Supplementary data

The *Dictyostelium discoideum* RNA-dependent RNA Polymerase RrpC Silences the Centromeric Retrotransposon DIRS-1 Post-transcriptionally and is Required for the Spreading of RNA Silencing Signals

Stephan Wiegand[1], Doreen Meier[2], Carsten Seehafer[1], Marek Malicki[1], Patrick Hofmann[1], Anika Schmith[3], Thomas Winckler[3], Balint Földesi[2], Benjamin Boesler[2], Wolfgang Nellen[2], Johan Reimegård[4], Max Käller[4], Jimmie Hällman[4], Olof Emanuelsson[4], Lotta Avesson[5], Fredrik Söderbom[6,7] and Christian Hammann[1,*]

[1]Ribogenetics@Biochemistry Lab, School of Engineering and Science, Molecular Life Sciences Research Center, Jacobs University Bremen, Campus Ring 1, DE-28759 Bremen, Germany
[2]Abteilung Genetik, Universität Kassel, Heinrich-Plett-Str. 40, DE-34132 Kassel
[3]Friedrich-Schiller-Universität Jena, Institut für Pharmazie, Lehrstuhl für Pharmazeutische Biologie, Semmelweisstraße 10, DE-07743 Jena
[4]KTH Royal Institute of Technology, Science for Life Laboratory (SciLifeLab Stockholm), School of Biotechnology, Division of Gene Technology, SE-171 65 Solna, Sweden
[5]Garvan Institute of Medical Research, 384 Victoria St Darlinghurst, NSW 2010, Australia
[6]Department of Cell and Molecular Biology, Biomedical Center, Uppsala University, Box 596, S-75124 Uppsala, Sweden and [7]Science for Life Laboratory, SE-75124 Uppsala, Sweden

* To whom correspondence should be addressed. Tel: +49-421-200-3247; Fax: +49-421-200-3249; Email: c.hammann@jacobs-university.de

This file contains Figures S1-S3, Tables S1-S5 and Supplementary Protocols and Supplementary References.

Item	**content**
Figure S1	Domain structure of the *D. discoideum* RNA-dependent RNA polymerases (RdRPs).
Figure S2	DIRS-1 strand specific qRT-PCR
Figure S3	Plasmid maps of ORF-GFP constructs
Table S1	DNA oligonucleotide sequences used in this study
Table S2	Ribo-probes and conditions for detection of strand-specific DIRS-1 transcripts by Northern blot Analyses
Table S3	PCR product as probe for Northern blot Analyses
Table S4	DNA Oligonucleotide probes and conditions for Northern blot Analyses.
Table S5	Relative DIRS-1 copy number in *rrpC*$^{-}$ cells

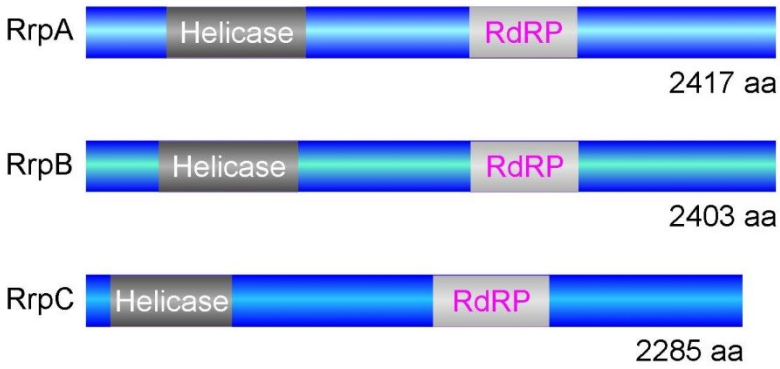


Figure S1. Domain structure of the *D. discoideum* RNA-dependent RNA polymerases (RdRPs). Each RdRP, RrpA, RrpB and RrpC, features an N-terminal Helicase domain and a central RdRP domain. The size of the proteins is shown in amino acids (aa).

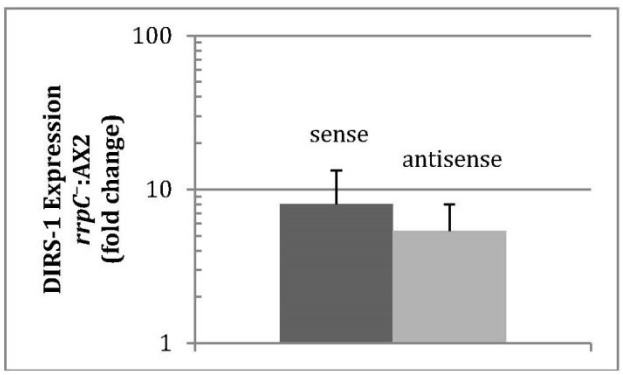


Figure S2. DIRS-1 strand specific qRT-PCR. Relative expression of DIRS-1 sense and antisense sequences in the *rrpC*⁻ strain relative to the AX2 wild type. Expression was monitored by using strand specific primers for the reverse transcription at position Q1 in the qRT-PCR and expression levels were normalized to the expression of GAPDH.

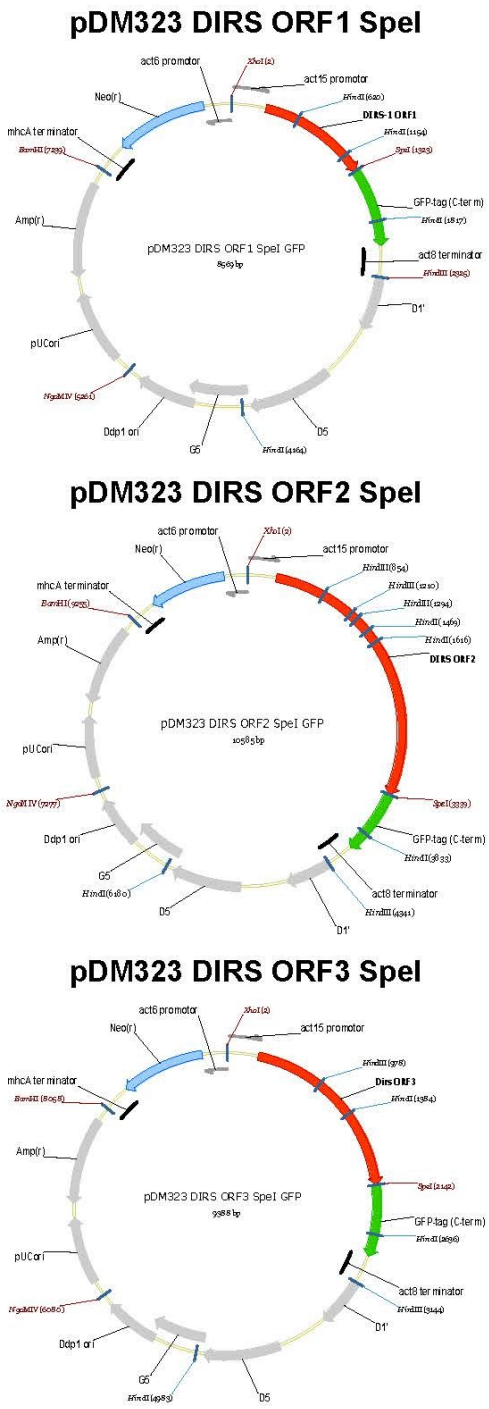


Figure S3: Plasmid maps of ORF-GFP constructs.

Table S1. DNA oligonucleotide sequences used in this study

Analysis	Oligo name	Sequence 5'-3'
qRT-PCR	Q-DIRS-01	GAATTAACATCGTTAGATGATACAAAC
	Q-DIRS-02	CAATACCTGACTTTGAGAGTGTTGATAG
	Q-DIRS-03	GTCCAATGTCAAGAAATTAGTTAACTC
	Q-DIRS-04	CAATTCTTTCCAAACTTGTTTGTGGTG
	Q-DIRS-05	GGTTTAGCAAGATCATCAGACTTGG
	Q-DIRS-06	CAGTATCATTTGATTTCCAACGACC
	Q-gpdA-01	GGTTGTCCCAATTGGTATTAATGG
	Q-gpdA-02	CCGTGGGTTGAATCATATTTGAAC
	Q-gpdA-03	GCGAAAACGTGGATTTTGTTACCATTTAC
Northern blotting	#465	GGCCATTACTCCCACTACTGG
	#66	ACCTCGATTGGAGTCAATGGA
	#156	CTATTCGAACAACTTTGGAA
	#157	GTTAATACGACTCACTATAGGGTCCATTGACTCCAATCG AGGT
	#158	GTTAATACGACTCACTATAGGGCTATTCGAACAACTTTG GAA
	#159	TCCATTGACTCCAATCGAGGT
	#160	AGTAAATCCAGTAGTGGGAGT
	#161	GTTAATACGACTCACTATAGGGTGATGCAATCTGATTTC GGA
	#162	GTTAATACGACTCACTATAGGGAGCAAATCCAGTAGTG GGAGT
	#163	GTGATGCAATCTGATTTCGGA
	#1198	TTAGCCTACGTCGCTCTCG
	#1199	GGGAACATAGTTGTACCACCT
	#1200	GGATGCCTGCCGGTTGCCCGGAGG
	#1647	GGCCAACAATTTTCTCAGCAAGAC
	#2642	GGCATGGCACTCTTGAAAAAG
Southern blotting	TRX-1for	GAACGAGCTCCATGGCCAATAGAGTAATTC
	TRX-1rev	CGCGGATCCTTATTTGTTTGCTTCTAGAGTA
FISH	DIRS1 forward	GGTAATCAAAACTTGGTCATTCC
	DIRS1 reverse	TCCATTGACTCCAATCGAGGT
Deep seq	RA3 adapter	TGGAATTCTCGGGTGCCAAGG
ORF-GFP fusions	ORFI forward	**GGATCCAAA**ATGTCTACCACTGTTAATAATAATGATG
	ORFI reverse	**ACTAGT**CTTCTTGTTCTTCTGAAAACGG
	ORFII forward	**GGATCCAAAATG**GCCGATCTAACAACTTC
	ORFII reverse	**ACTAGT**TGAAATTTGGACGACAGTATTTAAG
	ORFIII forward	**GGATCCAAAATG**CTAAACCCGATTCCGATTTC
	ORFIII reverse	**ACTAGT**CATGAGATTGGAAAGTTGAATAATC

Table S2. Ribo-probes and conditions for detection of strand-specific DIRS-1 transcripts by Northern blot Analyses

Ribo probe	Oligos for generation	Size of transcript	Hybridisation/ wash temperature	Exposition time
DIRS-1 LE s	#160, #161	312 nt	60°C / 65°C	≥ 1 day
DIRS-1 LE as	#162, #163	313 nt	60°C / 60°C	5 days
DIRS-1 RE s	#156, #157	305 nt	60°C / 65°C	≥ 1 day
DIRS-1 RE as	#158, #159	305 nt	60°C / 60°C	5 days

Table S3. PCR product as probe for Northern blot Analyses.

Target	Oligonucleotides for amplification	Size of fragment	Hybridisation/ wash temperature	Exposition time
actin6 mRNA	#1198 / #1199	275 bp	50°C / 54°C	≥ 1 Tag

Table S4. DNA Oligonucleotide probes and conditions for Northern blot Analyses.

Probe	Oligo	Hybridisation/ wash temperature	Exposition time
DIRS-1 LE siRNAs	#465	42°C / 42°C	≥ 1 day
DIRS-1 RE siRNAs	#66	42°C / 42°C	3 days
GFP siRNAs	#2642	42°C / 42°C	2-8 days
DdR-6	#1647	42°C / 42°C	≥ 8 h
U6	#1200	42°C / 42°C	≥ 6 h

Table S5. Phosphorimager readings for analysis of relative DIRS-1 copy numbers in *rrpC*⁻ cells

Digest	Strain	Thioredoxin Signal[a] (AU)	Loaded DNA[b] (%)	DIRS-1 Signal[a] (AU)	DIRS-1 signal (%)	Corrected DIRS-1 signal (%)
Eco RI	AX2	1068.77	100.0	4063.02	100.0	100
	rrpC⁻	1256.23	117.5	7259.90	178.7	152
Kpn I	AX2	858.13	100.0	2749.94	100.0	100
	rrpC⁻	875.36	102.0	4563.88	166.0	162

[a] Phosphorimager reading (Photostimulated luminescence) of bands using identical areas for related signals

[b] relative to AX2 wild type

Supplementary Material and Methods

Strand-specific qRT-PCR

For strand-specific qRT-PCR of DIRS-1 transcripts, cDNA was prepared using primers specific for DIRS-1 antisense (Q-DIRS-05) or sense RNA (Q-DIRS-06), and a GAPDH-specific primer (Q-gpdA-03). qRT-PCR was then performed from the cDNA preparations using primers Q-DIRS-01/-02 and Q-gpdA-01/02, respectively. All other procedures were as described in the main text experimental procedures.

Quantification of Southern blot signals

The membrane was incubated with stripping solution (0.4 M NaOH, 0.1% SDS) for 30 min at 45°C and two times for 10 min at room temperature. After neutralization with 2x SSC buffer, the membrane was again blocked with Church-Buffer and hybridized with a control DNA probe against the trx-1 gene, which was labeled with [α-^{32}P]dATP by random priming using the primers TRX-1for and TRX-1rev . All other procedures were performed as described in the main text experimental procedures.

9.3 Ergänzungen Manuskript 4

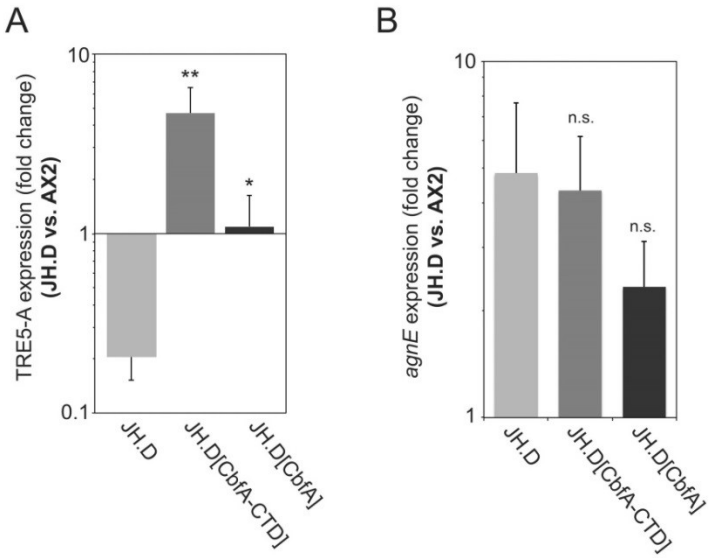
A host factor supports retrotransposition of the TRE5-A population in *Dictyostelium* cells by suppressing an Argonaute protein

---

Anika Schmith, Thomas Spaller, Åsa Fransson, Jonas Kjellin, Benjamin Boesler, Sandeep Ojha, Wolfgang Nellen, Christian Hammann, Fredrik Söderbom, Thomas Winckler

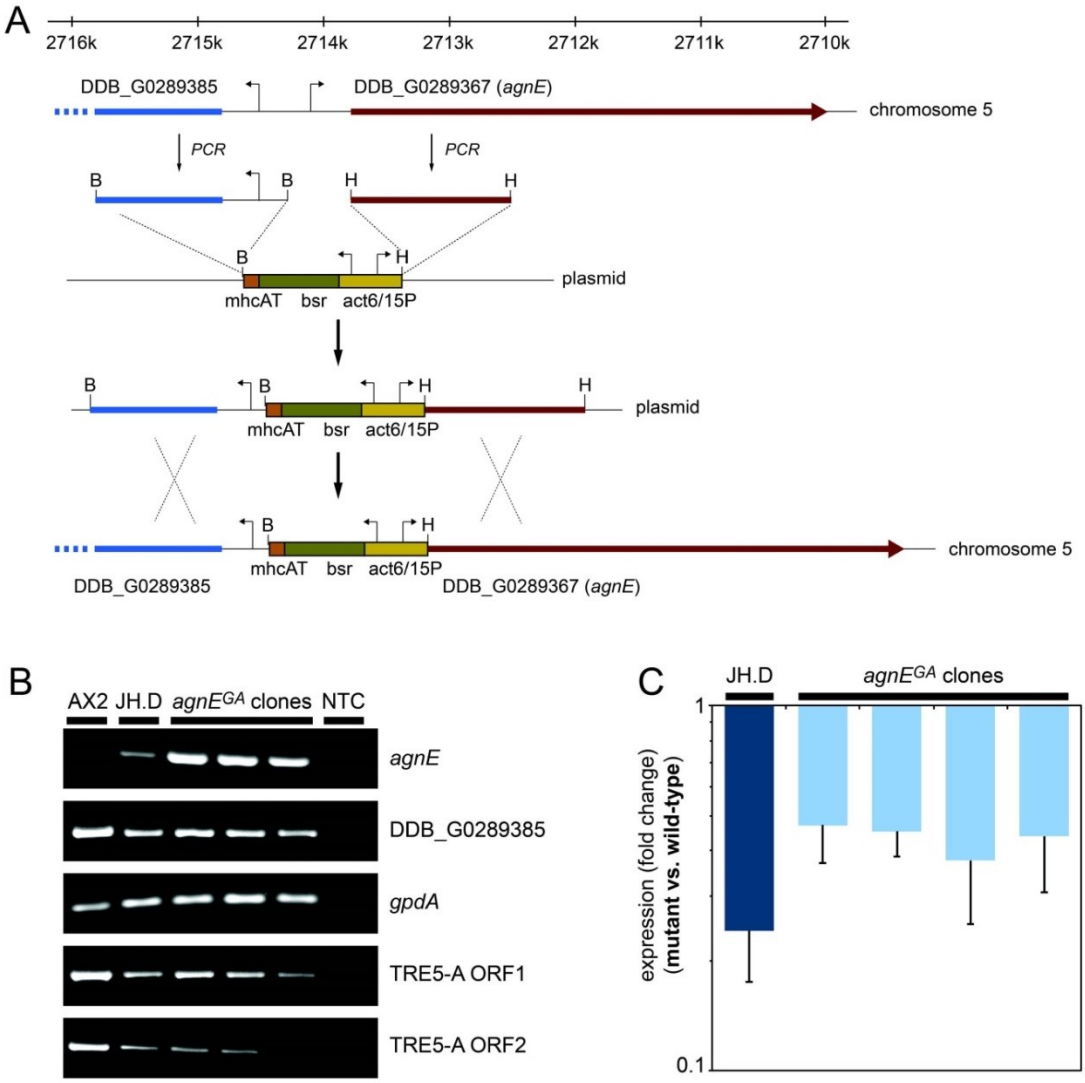
Publiziert in: *Mobile DNA, 2015*

Additional File 1: Figure S1: Functional complementation of strain JH.D with CbfA or its isolated carboxy-terminal domain.



RNA levels of retrotransposon TRE5-A (A) and *agnE* (B) were determined by qRT-PCR in JH.D cells, JH.D cells expressing full-length CbfA, and JH.D cells expressing CbfA-CTD. Expression levels in JH.D cells and JH.D transformants were compared to AX2 wild-type cells and are expressed as "fold change" of expression, meaning that values >1 represent overexpression of genes in the JH.D strains and a value of 1 would indicate complete reversion of the overexpression in JH.D cells. Values are means from six independent cultures ± SD. **p<0.01, relative to control AX2 cells (Student's t-test). Note that TRE5-A is actually overexpressed in the JH.D[CbfA-CTD] transformant, which is an effect of overexpression of CbfA-CTD (see Figure 2, main text).

Additional File 1: Figure S2: TRE5-A expression in *agnE*^{GA} mutants.



(A) Construction of *agnE* "gene activation" mutants. The *agnE* locus on chromosome 5 is indicated by nucleotide positions. The gene activation cassette consisted of a hybrid *actin6/actin15* promoter (arrows indicate transcription direction). The BamHI arm contained a 1070 bp DNA fragment covering part of the coding sequence of gene DDB_G0289385. The HindIII arm contained 1080 bp of *agnE* coding sequence, including the original translation start site. After double-recombination of the *agnE*^{GA} vector with genomic DNA, the expression of *agnE* was driven by the *act15* promoter, whereas expression of the neighboring gene DDB_G0289385 was unaffected. (B) Semi-quantitative RT-PCR analysis of RNA from AX2, JH.D, and three independent *agnE*^{GA} mutants demonstrating overexpression of *agnE*, normal expression of the neighboring gene DDB_G0289385 and *gpdA* (loading control), and silencing of TRE5-A (ORF1 and ORF2 sequences). NTC: no template control. (C) Quantitative RT-PCR of TRE5-A (ORF1) expression on RNA from JH.D and four *agnE*^{GA} mutants. Expression levels were compared to AX2 cells and are expressed as fold change of expression, meaning that values <1 represent lower levels of TRE5-A in the mutants relative to wild-type AX2 cells. Data represent means from four independent cultures of the indicated strains ± SD.

9.4 Plasmidliste

Plasmid-Name	Nummer	Inserts	wie kloniert?
pCR8/GW/TOPO -pyr56-ivi	20772	CbfA-Intron mit inversem Valin-tRNA-Gen (ivi)	puivi (11365) mit KpnI verdaut, ivi dann in 20692 eingefügt
pGEM7-GA-agnE	20769	agnE BamHI und Hind-Arm	agnE-Hind-Arm aus 20733 mit HindIII ausgeschnitten und über HindIII in 20737 eingefügt
pGEM7-GA-agnE-Bam-Arm	20737	agnE-Bam-Arm	mit BamHI aus 20734 ausgeschnitten und über BamHI in pGEM7-GA eingefügt
pCR8/GW/TOPO -pyr56-hyg(i)	20736	pyr56 mit hygromycin (hygromycin invers (i) zu pyr56)	20732 und 20731 mit BamHI verdaut, hygromycin dann in 20732 eingefügt
pCR8/GW/TOPO -pyr56-hyg	20735	pyr56 mit hygromycin (beide selbe Orientierung)	20732 und 20731 mit BamHI verdaut, hygromycin dann in 20732 eingefügt
pGEM-T-agnE-Bam-Arm	20734	agnE-Bam-Arm	PCR mit agnE-GA-01 und agnE-GA-02 an AX2 gDNA, Ligation in pGEM-T
pGEM-T-agnE-Hind-Arm	20733	agnE-Hind-Arm	PCR mit agnE-GA-03 und agnE-GA-04 an AX2 gDNA, Ligation in pGEM-T
pCR8/GW/TOPO -pyr56mut2	20732	pyr56	ds-Mutagenese an 20694 mit pyr56-KO-05 und pyr56-KO-06 durchgeführt
pGEM-T-hyg	20731	hygromycin	PCR mit bsr-ko-01 und bsr-ko-02 an pDM450 (Veltmann et al., 2008), Ligation in pGEM-T
pCR8/GW/TOPO -pyr56mut1	20694	pyr56	ds-Mutagenese an 20692 mit pyr56-KO-03 und pyr56-KO-04 durchgeführt

pUineu	20693	umps	Didi-tRNA-Gen (Val) mit EcoRI aus 11365 herausgeschnitten und religiert
pCR8/GW/TOPO -pyr56	20692	pyr56	PCR mit pyr56-KO-01 und pyr56-KO-02 an AX2 gDNA, Klonierung in pCR8/GW/TOPO
pGEM7-GA-agnC	20691	agnC Hind-Arm	agnC-Hind-Arm aus 20647 mit HindIII ausgeschnitten und über BamHI in 20660 eingefügt
pGEM7-GA-agnC-Bam-Arm	20660	agnC-Bam-Arm	mit BamHI aus 20646 ausgeschnitten und über BamHI in pGEM7-GA eingefügt
pGEM-T-agnC-Hind-Arm	20647	agnC03/04	PCR mit agnC-03 und agnC-04 an AX2 gDNA, dann Ligation in pGEM-T
pGEM-T-agnC-Bam-Arm	20646	agnC01/02	PCR mit agnC-01 und agnC-02 an AX2 gDNA, dann Ligation in pGEM-T
pUdelta	20645	deletiertes UMP-Synthase-Gen	11170 mit SnaBI und EcoRV verdaut, anschließend Religation
pDneo2a-N-Tap-agnE	20511	agnE	BamHI-Fragment, erhalten von AG Söderbom
pDneo2a-N-Tap-agnC	20510	agnC	BamHI-Fragment, erhalten von AG Söderbom
pET33b(+)-AChBP	20448	AChBP	Mit NcoI/BamHI aus 20416 herausgeschnitten, in pET33b(+) ligiert
pKOSG-IBA-dicty1-cbfBKO	20441	cbfB34/35 und cbfB36/37	als PCR-Produkte, Stargate Klonierung in pKOSG-IBA-dicty1
pET33b(+)-His-TEV-CTD	20425	His-TEV-CTD, CTD von *Dictyostelium discoideum*	mit NcoI/BamHI aus 20417 herausgeschnitten, in pET33b(+) ligiert

pDXA-GFP-CTD-Dfas	20424	CTD von *Dictyostelium fasciculatum*	mit BamHI aus 20395 heurausgeschnitten, in pDXA-GFP ligiert
pUC57-His-TEV-CTD	20417	CTD von *Dictyostelium discoideum*	Von GenScript käuflich erworben
pUC57-AChBP	20416	AChBP	Von GenScript käuflich erworben
pET33b-cbfBantigen	20415	cbfB	als NcoI/NotI-Fragment aus 20413 in pET-33b ligiert
pET15b-His-TEV-CTD	20414	CTD von *Dictyostelium discoideum*	mit KpnI aus 20409 herausgeschnitten, in pET15b-His-TEV (20369) ligiert
pGEM-cbfBantigen	20413	cbfB	PCR mit cbfB-32 und cbfB-33 an AX2 gDNA, dann Ligation in pGEM-T
pDXA-GFP2-CTD-Ppal	20412	CTD von *Polysphondylium pallidum*	mit BamHI aus 20404 herausgeschnitten, in pDXA-GFP2 (20368) ligiert
pBT-CTD-Dfas	20410	CTD von *Dictyostelium fasciculatum*	mit BamHI aus 20395 herausgeschnitten, in pBT (11670) ligiert
pGEM-CTD-TEV	20409	CTD-TEV	PCR mit CTD-TEV-01 und CTD-TEV-02 an 2.0231, dann Ligation in pGEM-T
pTOPO-cbfB(am)	20408	cbfB(am)	ds-Mutagenese an 20402 mit cbfB(am)-01 und cbfB(am)-02
pTOPO-cbfA-cDNADlac	20407	cbfA von *Dictyostelium lacteum*	PCR mit Dlac-01 und Dlac-05 an Dlac cDNA, Klonierung in pCR-XL-TOPO (XL PCR Cloning Kit von Invitrogen)
pTOPO-cbfA-Dlac	20406	cbfA von *Dictyostelium lacteum*	PCR mit Dlac-01 und Dlac-02 an Dlac gDNA, anschließend Klonierung in pCR-XL-TOPO (XL PCR Cloning Kit von Invitrogen)

pBT-CTD-Ppal	20404	CTD von *Polysphondylium pallidum*	mit BamHI aus 20401 herausgeschnitten, in pBT (11670) ligiert
pDXA-GFP2-CTD-Ddis	20403	CTD von *Dictyostelium discoideum*	mit BamHI aus 20400 herausgeschnitten, in pDXA-GFP2 (20368) ligiert
pTOPO-cbfBKI	20402	CbfB	PCR mit cbfB-KI-01 und cbfB-KI-02 an AX2 gDNA, Klonierung in pCR-XL-TOPO (XL PCR Cloning Kit von Invitrogen)
pGEM_CTD_Ppal	20401	CTD von *Polysphondylium pallidum*	PCR mit GFP-PpalCTD-01 und GFP-PpalCTD-02 an Pp gDNA dann Ligation in pGEM-T
pBT-CTD-Ddis	20400	CTD von *Dictyostelium discoideum*	mit BamHI aus 20396 herausgeschnitten, in pBT ligiert
pGEM_CTD_Ddis	20396	CTD von *Dictyostelium discoideum*	PCR mit GFP-DdisCTD-01 und V-02 an 2.0231, dann Ligation in pGEM-T
pGEM_CTD_Dfas	20395	CTD von *Dictyostelium fasciculatum*	PCR mit GFP-DfasCTD-01 und GFP-DfasCTD-02 an Dfas gDNA, Ligation in pGEM-T
pGEM-CTD_Ppal_mut	20379	CTD von *Polysphondylium pallidum* (enthält Mutationen)	PCR mit GFP-PpalCTD-01 und GFP-PpalCTD-02 an Ppal gDNA von Ppla PN500, Ligation in pGEM-T
pGEM-CTD_Dfas_mut	20378	CTD von *Dictyostelium fasciculatum* (enthält Mutationen)	PCR mit GFP-DfasCTD-01 und GFP-DfasCTD-02 an Dfas gDNA, Ligation in pGEM-T
pGEM-CTD_Ddis_mut	20377	CTD von *Dictyostelium discoideum* (enthält Mutationen)	PCR mit GFP-DdisCTD-01 und GFP-DdisCTD-02 an 2.0231, Ligation in pGEM-T

10. Danksagung

Angefertigt wurde diese Arbeit am Lehrstuhl für Pharmazeutische Biologie der Friedrich-Schiller-Universität Jena unter Anleitung von Prof. Dr. Thomas Winckler. Ich danke ihm für die Möglichkeit, dieses sehr interessante Themengebiet während meiner Promotion untersuchen zu können. Auch danke ich ihm für die ständige Bereitschaft, offene Fragen und Probleme rund um diese Arbeit stets kompetent zu beantworten oder zu diskutieren.

Dr. Marco Groth vom Institut für Altersforschung Jena danke ich für die Durchführung der RNA-Seq-Experimente, deren Ergebnisse meine Arbeit einen großen Schritt voranbringen konnten. Prof. Dr. Fredrik Söderbom, Universität Uppsala, Prof. Dr. Christian Hammann, Universität Bremen, und Prof. Dr. Wolfgang Nellen, Universität Kassel, danke ich für die Überlassung von Stämmen und Plasmiden und ihre Anmerkungen und Empfehlungen rund um das Thema RNA-Interferenz in *Dictyostelium discoideum*. Ebenfalls danke ich Prof. Dr. Gernot Glöckner, Universität Köln, für die Zurverfügungstellung von Genomdaten vor deren Veröffentlichung.

Ein großes Dankeschön verdienen alle derzeitigen und ehemaligen Mitarbeiter des Lehrstuhls. Ihr seid ein so herzliches, freundliches und hilfsbereites Team, mit dem es eine Freude war, sich dem Laboralltag mit allen seinen Tücken, Anforderungen und auch Erfolgen zu stellen. Potsdam, Teamlauf, Sommerpraktikum, Holunderblüten, „Braten“ und vieles mehr sind Erfahrungen und Erlebnisse mit euch, die ich nicht mehr vergessen werde.

Selbstverständlich danke ich meinen Eltern und meiner Schwester. Durch ihre Unterstützung und Motivation trugen und tragen auch sie einen großen Beitrag zum Gelingen dieser Arbeit bei. Meiner eigenen kleinen Familie danke ich für ihre Liebe, Freude und Kraft. Danke, dass ich euch habe.

11. Eigenständigkeitserklärung

Hiermit erkläre ich, gemäß der Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena vom 4. Dezember 2012, dass die vorliegende Arbeit selbstständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet wurden.

Die geltende Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena ist mir bekannt.

Die Hilfe eines Promotionsberaters wurde von mir nicht in Anspruch genommen.

Niemand hat von mir mittelbar oder unmittelbar geldwerte Leistungen für Arbeiten bekommen, die im Zusammenhang mit der vorgelegten Arbeit stehen.

Die vorliegende Dissertation wurde weder bei einer anderen Prüfungsbehörde noch einer anderen Hochschule in gleicher oder ähnlicher Form eingereicht.

Müllheim, den 23.10.2015

Anika Schmith

12. Liste der wissenschaftlichen Veröffentlichungen

Bilzer A., Dölz H., Reinhardt A., **Schmith A.**, Siol O., Winckler T. (2011) The C-module-binding factor supports amplification of TRE5-A retrotransposons in the *Dictyostelium discoideum* genome. *Eukaryot. Cell.* 10(1): 81-6.

Thewes S., Krohn S., **Schmith A.**, Herzog S., Stach T., Weissenmayer B., Mutzel R. (2012) The calcineurin dependent transcription factor TacA is involved in development and the stress response of *Dictyostelium discoideum*. *Eur J Cell Biol*. (10):789-799.

Schmith A., Groth M., Ratka J., Gatz S., Spaller T., Siol O., Glöckner G., Winckler T. (2013) Conserved gene regulatory function of the carboxy-terminal domain of dictyostelid C-module-binding factor. *Eukaryot. Cell*. 12(3): 460-468.

Wiegand S., Meier D., Seehafer C., Malicki M., Hofmann P., **Schmith A.**, Winckler T., Földesi B., Boesler B., Nellen W., Reimegard J., Käller M., Hällman J., Emanuelsson O., Avesson L., Söderbom F., Hammann C. (2014) The *Dictyostelium discoideum* RNA-dependent RNA polymerase RrpC silences the centromeric retrotransposon DIRS-1 post-transcriptionally and is required for the spreading of RNA silencing signals. *Nucleic Acids Res*. 42(5): 3330-3345.

Schmith A., Spaller T., Gaube F., Fransson Å., Boesler B., Ojha S., Nellen W., Hammann C., Söderbom F., Winckler T. (2015) A host factor supports retrotransposition of the TRE5-A population in *Dictyostelium* cells by suppressing an Argonaute protein. *Mobile DNA.* 6: 14.

Schmith A., Fransson Å., Boesler B., Meier D., Wiegand S., Nellen W., Hammann C., Söderbom F., Winckler T. (2014) Retrotransposition of TRE5-A in the *Dictyostelium discoideum* genome is sustained by suppression of RNA interference by a host encoded factor. Posterpräsentation Annual International *Dictyostelium* Conference 2014, Potsdam.

Malicki M., Villafano G., Wiegand S., Söderbom F., Spaller T., **Schmith A.** ,Winckler T., Hammann C. (2014) Skipper retrotransposon mobilization in *Dictyostelium discoideum*. Poster Annual International *Dictyostelium* Conference 2014, Potsdam.